Asymptotic properties of QML estimators for VARMA models with time-dependent coefficients: Part I

Abdelkamel Alj *, Christophe Ley† and Guy Mélard‡

June 5, 2015

Abstract

This paper is about vector autoregressive-moving average (VARMA) models with time-dependent coefficients to represent non-stationary time series. Contrarily to other papers in the univariate case, the coefficients depend on time but not on the length of the series n. Under appropriate assumptions, it is shown that a Gaussian quasi-maximum likelihood estimator is almost surely consistent and asymptotically normal. The theoretical results are illustrated by means of two examples of bivariate processes. It is shown that the assumptions underlying the theoretical results apply. In the second example the innovations are also marginally heteroscedastic with a correlation ranging from -0.8 to 0.8. In the two examples, the asymptotic information matrix is obtained in the Gaussian case. Finally, the finite-sample behaviour is checked via a Monte Carlo simulation study for n going from 25 to 400. The results confirm the validity of the asymptotic properties even for short series and reveal that the asymptotic information matrix deduced from the theory is correct.

Key words and phrases : Non-stationary process; multivariate time series; time-varying models.

Running title: Asymptotics of QMLEs for tdVARMA models

*Université Moulay Ismail, Faculté des Sciences juridiques, économiques et sociales, B.P. 3102 Toulal, Meknès, Morocco. This work was started while at Université libre de Bruxelles, Faculty of Sciences, Dept of Mathematics, Blv du Triomphe CP 210, B-1050 Brussels, Belgium (e-mail: abdelkamel.alj@gmail.com).

†Université libre de Bruxelles, Faculty of Sciences, Dept of Mathematics, Blv du Triomphe CP 210, B-1050 Brussels, Belgium (e-mail: chrisley@ulb.ac.be).

‡Université libre de Bruxelles, SBS-EM, ECARES, avenue Franklin Roosevelt 50 CP 114/04, B-1050 Brussels, Belgium (e-mail: gmelard@ulb.ac.be).

1 Introduction

A large part of the literature on time series models is concerned with stationary models. This is of course due to the ensuing mathematical simplifications of stationarity. Even in that simple context, an asymptotic analysis is not necessarily easy; indeed, the celebrated autoregressive-moving average (ARMA) models popularized by Box and Jenkins (Box et al., 2008) require several pages in Brockwell & Davis (1991, pp. 375-396) for the derivation of their asymptotic properties. However, the assumption of invariance over time (especially for long time intervals) is difficult to justify in most practical situations. Therefore, recent years have seen an increasing interest in models with time-dependent coefficients and non-stationary time series. Initiated by the seminal work Quenouille (1957), models with time-dependent or time-varying coefficients for univariate time series have been investigated over the years by, *inter alia*, Whittle (1965), Subba Rao (1970), Tjøstheim (1984), Kwoun & Yajima (1986), Singh & Peiris (1987), Priestley (1988), Grillenzoni (1990), Dahlhaus (1996a, b, c, 1997), Bibi & Francq (2003), Azrak & Mélard (2006) and Triantafyllopoulos & Nason (2007). We refer to the introduction of Azrak & Mélard (2006) or of Van Bellegem & Dahlhaus (2006) for further references. In several of these papers, the coefficients of the ARMA models are not constant but are deterministic functions of time. Also the innovation variance can be a deterministic function of time instead of being constant, such as in Van Bellegem & von Sachs (2004). We can speak of marginal heteroscedasticity by opposition to conditional heteroscedasticity which is encountered in ARCH and GARCH models. All these functions of time are supposed to depend on a small number of parameters. Other somewhat related recent approaches include generalized autoregressive score (GAS) models of Creal et al. (2013) and testing parameter constancy against deterministically time-varying parameters, e.g. Teräsvirta et al (2010, Section 6.3) and references therein, generalized to VAR models in Teräsvirta & Yang (2014).

The present paper inscribes itself in this line of research but for multivariate time series. The generality of the models we consider evidently entails numerous challenges, since the convenient asymptotic theory of stationary ergodic processes does no longer apply. Also, the asymptotic theory of time series models makes a large use of Fourier transforms and, consequently, of what is called spectral analysis. As a consequence, deriving conditions for consistency and asymptotic normality of estimators of the coefficients, as well as obtaining the asymptotic covariance matrix, becomes highly complicated.

We consider multidimensional time series models, with particular emphasis on *vector ARMA* (VARMA) models, the multivariate extension of the ARMA models. The main difference between ARMA and VARMA models lies in the fact that the coefficients change from scalars to squared matrices. The main developments in the area of statistical inference of standard stationary VARMA models are due to Kohn (1978), Hannan & Deistler (1988), Francq & Raïssi (2007) and, quite recently, Boubacar Mainassara & Francq (2011) who study the consistency and asymptotic normality of quasi-maximum likelihood estimators for weak VARMA models. However, the field of *time-dependent VARMA (tdVARMA)* models with marginally heteroscedastic innovation covariance matrix remains largely unexplored. An exception is Dahlhaus (2000) using an entirely different approach and assuming that the coefficients depend on time t but also on the length of the series n trough their ratio t/n. Here we assume dependency on t only. Even if our theory is illustrated on pure VAR examples, it should be emphasized that it is valid for VMA and VARMA models, like in Dahlhaus (2000). Note that Lütkepohl (2005, Chap. 14) treats tdVAR models by Gaussian maximum likelihood but does not discuss asymptotic properties in the general case.

Thus we want to fill in this gap in the literature by extending to the multivariate setting the methodology of Azrak & Mélard (2006) who, to the best of the authors' knowledge, were the first to obtain asymptotic properties of estimators for the general class of univariate time-dependent ARMA models by having recourse to *quasi-maximum likelihood estimation (QMLE)*.

Like other QMLE approaches, the estimation method in Azrak & Mélard (2006) does not use the true, unknown, density of the observations but rather acts as if that density were Gaussian, thus using the Gaussian log-likelihood, which is an extension of the generalized least-squares method since it takes care of possible heteroscedasticity. There is no assumption of stationarity but, although it is not illustrated in our examples, there is an adjustment in the asymptotic theory for allowing non-normal observations. One major advantage of QMLE is that the Gaussian likelihood function can be computed exactly, with an efficient algorithm, Alj et al. (2015c), and this is very important for short time series. The main task in the Azrak-Mélard approach, hence also in our extension, consists in checking conditions from two crucial theorems in Klimko & Nelson (1978), which respectively ensure existence of an almost surely (a.s.) consistent estimator and prove asymptotic normality of that estimator, whilst providing the asymptotic covariance matrix. This is precisely what we are aiming at but, as we shall see in the rest of this paper, it is all but an easy task.

Let us briefly comment on two other univariate approaches, and explain why we have not opted for extending those.

- The Dahlhaus approach (Dahlhaus 1996a, b, c, 1997) has allowed to obtain asymptotic results for a class of locally stationary processes including heteroscedastic ARMA processes with time-dependent coefficients. Dahlhaus uses either a spectral-based or a maximum likelihood estimation method. His asymptotics are based on rescaling time, i.e. t/n; therefore this is not related to our approach where there is no such requirement. Moreover local stationarity implies that the coefficients are continuous functions of time (and even two time differentiable functions), which is not necessarily the case here. Azrak & Mélard (2011) show a univariate example where Azrak & Mélard (2006)'s theory holds but the assumptions of local stationarity are not valid. Note also that Dahlhaus (2000)'s theory for multivariate processes assumes a Gaussian process whereas we assume only existence of 8th-order moments.

- The Bibi-Francq approach (Bibi & Francq 2003) applies the quasi-least squares estimation method and gives asymptotic results for cyclical ARMA models with non constant periods. Although only 4th-order moments are needed, the theory was not developed for multivariate processes until now.

Another related approach, see for instance Tiao & Grupe (1980) or Basawa & Lund (2001), consists in ARMA models with coefficients that vary as periodic functions of time, see also Hindrayanto et al. (2010). If the period s is an integer, s consecutive variables can be stacked as a vector which satisfies a stationary VARMA model. Here we do not assume periodic coefficients although, to simplify the derivations, our two examples will have periodic coefficients or innovation covariance matrix, but with large or irrational periods. Therefore stacking the variables will be practically inoperative and standard asymptotic theory for stationary VARMA models will not apply.

Most of the technical lemmas used in the present paper and their proofs are given in a technical appendix denoted here 'TA', see Alj et al. (2015a). The second part of this paper (Alj et al., 2015b) will deal with the more general case where the coefficients of the model depend on time t but also possibly on the number of observations n of the series. However, even in that case, the theory differs from Dahlhaus' approach in the sense that the coefficients do not need to be continuous functions of time. In the univariate case, an example is provided by Azrak & Mélard (2011). The technical appendix, Alj et al. (2015a), will be shared by Alj et al. (2015b). The reason

to separate the material in two parts is that while theorems related to martingale sequences are enough in the context of the present paper, the second part requires martingale arrays, although the technicalities are more or less the same.

Another aspect of the present paper is that it provides an alternative theory for the asymptotics of standard VARMA models that does not rely on stationarity or ergodicity arguments, although in the standard case, the assumptions will imply the usual conditions on the roots of the autoregressive and moving average polynomials in the lag operator. Our alternative theory also avoids spectral analysis.

The paper is organized as follows. In Section 2, we first develop asymptotics for quasi-maximum likelihood estimators in a general multivariate time series model which is not necessarily stationary. Then, in Section 3, we focus our attention on tdVARMA models: after setting the notations, we analyze pure VAR and pure VMA representations, with an illustration, before finally stating the main theorem for the tdVARMA case. We illustrate our theoretical findings by means of two examples in Section 4, and examine the finite-sample behavior of our estimators via a Monte Carlo simulation study in Section 5. Finally, Appendix A collects the main proofs and Appendix B contains a verification of the main assumptions for the two examples studied with a few nice mathematical derivations.

2 QMLE for a general multivariate time series model

2.1 Some preliminaries

Let $\{x_t : t \in \mathbf{N}\}$ be a stochastic process defined on a probability space (Ω, F, P_θ), taking values in $\mathbf{R}^r$, and whose distribution depends on a vector $\theta = (\theta_1, ..., \theta_m)^T$ of unknown parameters to be estimated, with θ lying in some open set Θ of a Euclidean space $\mathbf{R}^m$. Let $E_\theta(.)$ and $E_\theta(./.)$ denote expectation and conditional expectation under P_θ, respectively. The true value of θ is denoted by $\theta^0 = (\theta^0_1, ..., \theta^0_m)^T$, assumed to be an interior point of Θ. Let $\{F_t : t \in \mathbf{N}\}$ be an increasing sequence of sub-sigma algebras of F with F_t generated by $\{x_u : u = 1, 2, ..., t\}$ with $F_0 = \{\emptyset, \Omega\}$ such that, for each t, x_t is measurable with respect to F_t. Given a set of observations $\{x_t : t = 1, 2, ..., n\}$, we want to estimate θ by trying to minimize the general real-valued objective function $Q_n(\theta) = Q_n(\theta; x_1, ..., x_n)$ which depends on θ and the observations $\{x_t : t = 1, 2, ..., n\}$. Therefore we solve the system of equations

$$\frac{\partial Q_n(\theta)}{\partial \theta_i} = 0 \quad \text{for} \quad i = 1, ..., m.$$

We suppose that the objective function $Q_n(\theta)$ is twice continuously differentiable in Θ. Let $\widehat{\theta}_n = (\widehat{\theta}_1, ..., \widehat{\theta}_m)^T$ be a sequence of estimators indexed by n. Klimko & Nelson (1978) showed conditions for strong consistency and asymptotic normality of $\widehat{\theta}_n$, see also Hall & Heyde (1980, pp. 174-176) and Taniguchi & Kakizawa (2000, pp. 97-98).

2.2 General theory of quasi-maximum likelihood estimation

Denote

$$e_t(\theta) = x_t - \hat{x}_{t/t-1}(\theta) \quad \text{with} \quad \hat{x}_{t/t-1}(\theta) = E_\theta(x_t/F_{t-1}), \tag{2.1}$$

for which obviously $E_\theta(e_t(\theta)) = 0$. We denote by

$$\Sigma_t(\theta) = E_\theta \left[e_t(\theta) e_t^T(\theta)/F_{t-1} \right]$$

the conditional covariance matrix given F_{t-1}. The quasi-likelihood function $L_n(\theta; x_1, ..., x_n)$ computed as if the process were Gaussian is given by

$$L_n(\theta; x_1, ..., x_n) \quad = \quad (2\pi)^{-nr/2} \prod_{t=1}^{n} \det\left(\Sigma_t(\theta)\right)^{-1/2} \exp\left\{ -\frac{1}{2} \sum_{t=1}^{n} e_t^T(\theta) \Sigma_t^{-1}(\theta) e_t(\theta) \right\}.$$

We take the objective function

$$Q_n(\theta) \quad = \quad -\log\left(L_n(\theta; x_1, ..., x_n)\right) = \frac{1}{2} \sum_{t=1}^{n} \alpha_t(\theta) + \frac{rn}{2} \log(2\pi), \tag{2.2}$$

with

$$\alpha_t(\theta) = \log\left(\det\left(\Sigma_t(\theta)\right)\right) + e_t^T(\theta) \Sigma_t^{-1}(\theta) e_t(\theta).$$

Then the QMLE of θ is defined as any measurable solution $\widehat{\theta}_n$ of

$$\text{arg}\ \min_{\theta \in \Theta} \quad Q_n(\theta). \tag{2.3}$$

In order to check the assumptions of the Klimko & Nelson (1978) theorems, we proceed like Azrak & Mélard (2006) and we make some additional assumptions as follows. Let the r-vector stochastic process $\{x_t : t \in \mathbf{N}\}$ be such that $E_\theta\left(\|x_t\|^2\right) < \infty$ for all θ and $e_t(\theta)$ and $\Sigma_t(\theta)$ are almost surely twice continuously differentiable in Θ. Henceforth, for simplicity, we denote $\left[E_\theta\{.(\theta)\}\right]_{\theta=\theta^0}$ by $E_{\theta^0}\{.(\theta)\}$. We suppose that there exist two positive constants C_1 and C_2 such that for all $t \geq 1$:

H$_{2.1}$ $\quad E_{\theta^0}\left\{\left|\frac{\partial \alpha_t(\theta)}{\partial \theta_i}\right|^4\right\} \leq C_1$ for $i = 1, ..., m$;

H$_{2.2}$ $\quad E_{\theta^0}\left\{\left|\frac{\partial^2 \alpha_t(\theta)}{\partial \theta_i \partial \theta_j} - E_\theta\left(\frac{\partial^2 \alpha_t(\theta)}{\partial \theta_i \partial \theta_j}/F_{t-1}\right)\right|^2\right\} \leq C_2$ for $i, j = 1, ..., m$.

Suppose further that

H$_{2.3}$

$$\lim_{n\to\infty} \frac{1}{2n} \sum_{t=1}^{n} E_{\theta^0}\left\{\frac{\partial^2 \alpha_t(\theta)}{\partial \theta_i \partial \theta_j}/F_{t-1}\right\} = V_{ij} \quad \text{a.s.} \quad \text{for} \quad i, j = 1, ..., m,$$

where $V = (V_{ij})_{1\leq i,j\leq m}$ is a strictly positive definite matrix of constants;

H$_{2.4}$

$$\lim_{n\to\infty} \sup_{\Delta\downarrow 0} (n\Delta)^{-1} \left|\sum_{t=1}^{n}\left(\left\{\frac{\partial^2 \alpha_t(\theta)}{\partial \theta_i \partial \theta_j}\right\}_{\theta=\theta^*} - \left\{\frac{\partial^2 \alpha_t(\theta)}{\partial \theta_i \partial \theta_j}\right\}_{\theta=\theta^0}\right)\right| < \infty \quad \text{a.s.}$$

for $i, j = 1, ..., m$, where θ^* is a point of the straight line joining θ^0 to every θ, such that $\|\theta - \theta^0\| < \Delta$, $0 < \Delta$, where $\|.\|$ is the Euclidean norm.

The main notations and assumptions being settled, we are ready to state the two main theorems of the present paper.

Theorem 2.1 *Suppose that Assumptions* **H**$_{2.1}$ $-$ **H**$_{2.4}$ *hold. Then there exists a sequence of estimators* $\widehat{\theta}_n = (\widehat{\theta}_1, ..., \widehat{\theta}_m)^T$ *such that* $\widehat{\theta}_n \to \theta^0$ *a.s. and, for any* $\epsilon > 0$, *there exists an event* E *with* $P_{\theta^0}(E) > 1-\epsilon$ *and an* n_0 *such that on* E, *for any* $n > n_0$, $\{\partial Q_n(\theta)/\partial \theta_i\}_{\theta=\widehat{\theta}_n} = 0$, *for* $i = 1, 2, ..., m$, *and* $Q_n(\theta)$ *attains a relative minimum at* $\widehat{\theta}_n$.

The proof of Theorem 2.1 is given in Appendix A.

Theorem 2.2 *If the assumptions* **H**$_{2.1}$ $-$ **H**$_{2.4}$ *are satisfied, as well as* **H**$_{2.5}$

$$\frac{1}{n}\sum_{t=1}^{n} E_{\theta^0}\left(\frac{\partial \alpha_t(\theta)}{\partial \theta}\frac{\partial \alpha_t(\theta)}{\partial \theta^T}/F_{t-1}\right) - \frac{1}{n}\sum_{t=1}^{n} E_{\theta^0}\left(\frac{\partial \alpha_t(\theta)}{\partial \theta}\frac{\partial \alpha_t(\theta)}{\partial \theta^T}\right) \xrightarrow{a.s.} 0 \quad \textit{as} \quad n \to \infty,$$

then

$$n^{1/2}(\widehat{\theta}_n - \theta^0) \xrightarrow{L} \mathcal{N}(0, V^{-1}WV^{-1}),$$

where $\xrightarrow{L}$ *indicates convergence in law and* $W = (W_{ij})_{1\leq i,j\leq m}$ *is a strictly positive*

definite matrix defined by

$$W = \lim_{n\to\infty} \frac{1}{4n} \sum_{t=1}^{n} E_{\theta^0} \left(\frac{\partial \alpha_t(\theta)}{\partial \theta} \frac{\partial \alpha_t(\theta)}{\partial \theta^T} \right).$$

For the proof of Theorem 2.2 we use the Central Limit Theorem for martingale differences of Basawa & Prakasa Rao (1980).

3 VARMA models with time-dependent coefficients

3.1 tdVARMA models: definition and notations

The process $\{x_t : t \in \mathbf{N}\}$ is called a zero mean r-vector mixed autoregressive-moving average process of order $(p,\ q)$ with time-dependent coefficients, and is denoted by tdVARMA $(p,\ q)$, if and only if it satisfies the equation

$$x_t = \sum_{i=1}^{p} A_{ti} x_{t-i} + g_t \epsilon_t + \sum_{j=1}^{q} B_{tj} g_{t-j} \epsilon_{t-j}, \tag{3.1}$$

where p and q are integer constants, $\{\epsilon_t : t \in \mathbf{N}\}$ is an independent white noise process, consisting of independent random variables, not necessarily identically distributed, with zero mean, covariance matrix Σ which is invertible, and finite fourth-order moments, and where the coefficients A_{t1}, ..., A_{tp} and $B_{t1}, ..., B_{tq}$, as well as the $r \times r$ matrix g_t, are deterministic functions of time t. The initial values $x_t, t < 1$, and ϵ_t, $t < 1$, are supposed to be equal to the zero vector. In the sequel, we will also use $A_{t0} = B_{t0} = I_r$ with I_r the r-dimensional identity matrix and set to zero the coefficients A_{tk} with $k > p$ and B_{tk} with $k > q$, for all t. Writing $\otimes$ the Kronecker product, we let

$$\kappa_t = E\left(\mathrm{vec}(\epsilon_t \epsilon_t^T)\,\mathrm{vec}(\epsilon_t \epsilon_t^T)^T\right) = E\left((\epsilon_t \epsilon_t^T) \otimes (\epsilon_t \epsilon_t^T)\right),$$

which depends on t, in general. For $k, l \in \mathbf{N}$ and $k \neq l$, we consider the matrix

$$E\left(\mathrm{vec}(\epsilon_{t-k} \epsilon_{t-l}^T)\,\mathrm{vec}(\epsilon_{t-l} \epsilon_{t-k}^T)^T\right) = K_{r,r}(\Sigma \otimes \Sigma),$$

which does not depend on t, k or l and where the $r^2 \times r^2$ matrix $K_{r,r}$ is the *commutation matrix*. See Kollo & von Rosen (2005, p. 79) or TA Lemma 4.2.

Let us now consider the parametric model corresponding to (3.1), namely

$$x_t = \sum_{i=1}^{p} A_{ti}(\theta)x_{t-i} + e_t(\theta) + \sum_{j=1}^{q} B_{tj}(\theta)e_{t-j}(\theta), \tag{3.2}$$

where the $e_t(\theta)$ can be considered as the residuals of the model and are defined as in (2.1) and where $A_{ti}(\theta)$, for $i = 1, ..., p$, and $B_{tj}(\theta)$, for $j = 1, ..., q$, are the parametric coefficients. Furthermore the covariance matrix of $e_t(\theta)$ is parametrized as

$$\Sigma_t(\theta) =^{\text{def}} E_\theta\left(e_t(\theta)e_t^T(\theta)\right) = g_t(\theta)\Sigma g_t^T(\theta).$$

For $\theta = \theta^0$, we have $A_{ti}(\theta^0) = A_{ti}$, $B_{tj}(\theta^0) = B_{tj}$, $g_t(\theta^0) = g_t$, $e_t(\theta^0) = g_t\epsilon_t$ and $\Sigma_t =^{\text{def}} \Sigma_t(\theta^0) = E\left(e_t(\theta^0)e_t^T(\theta^0)\right) = g_t\Sigma g_t^T$. We assume that the m-dimensional vector θ contains all the parameters of interest to be estimated, those in the coefficients $A_{t1}(\theta), ..., A_{tp}(\theta)$, $B_{t1}(\theta), ..., B_{tq}(\theta)$ and $g_t(\theta)$ but not the nuisance parameters in the scale factor matrix Σ which are estimated separately. In usual VARMA(p, q) models, the coefficients $A_1(\theta), ..., A_p(\theta)$, $B_1(\theta), ..., B_q(\theta)$ and $g_t(\theta)$ do not depend on t, and the parameters are the coefficients themselves. Note that for a given θ we have

$$\hat{x}_{t/t-1}(\theta) \quad = \quad E_\theta(x_t/F_{t-1}) = \sum_{i=1}^{p} A_{ti}(\theta)x_{t-i} + \sum_{j=1}^{q} B_{tj}(\theta)e_{t-j}(\theta).$$

According to the assumptions made about initial values, it is possible to write out properly the pure autoregressive and the pure moving average representation of the model (3.2), as we shall see in the next section.

3.2 The pure autoregressive and the pure moving average representations

By using the assumption about initial values and using (3.2) recurrently, Mélard (1985) and Azrak & Mélard (2015) have established expressions for the pure autoregressive representation and the pure moving average representation of tdVARMA processes. In the univariate case these representations can be found in Azrak & Mélard (2006). In our setting, for any θ the pure autoregressive representation corresponds to

$$x_t = \sum_{k=1}^{t-1} \pi_{tk}(\theta)x_{t-k} + e_t(\theta), \tag{3.3}$$

where the coefficients $\pi_{tk}(\theta)$ can be obtained from the autoregressive and moving average coefficients by using the following recurrences (see Mélard, 1985, pp. 43-45):

$$\begin{aligned}
\pi_{t0}^{(0)}(\theta) &= I_r, \quad \pi_{tj}^{(0)}(\theta) = A_{tj}(\theta), \quad \widetilde{\pi}_{tj}^{(0)}(\theta) = B_{tj}(\theta), \quad \text{for} \quad j = 1, ..., t-1, \\
\pi_{tj}^{(k)}(\theta) &= \pi_{tj}^{(k-1)}(\theta) - \widetilde{\pi}_{tk}^{(k-1)}(\theta) A_{t-k,j-k}(\theta), \quad \text{for} \quad j = k, ..., t-1, \\
\widetilde{\pi}_{tj}^{(k)}(\theta) &= \widetilde{\pi}_{tj}^{(k-1)}(\theta) - \widetilde{\pi}_{tk}^{(k-1)}(\theta) B_{t-k,j-k}(\theta), \quad \text{for} \quad j = k+1, ..., t-1,
\end{aligned}$$

and $\pi_{tk}(\theta) = \pi_{tk}^{(k)}(\theta)$ for $k = 1, ..., t-1$. By (3.3) we of course have $e_t(\theta) = x_t - \sum_{k=1}^{t-1} \pi_{tk}(\theta) x_{t-k}$, and consequently its first three derivatives with respect to θ are given by

$$\begin{aligned}
\frac{\partial e_t(\theta)}{\partial \theta_i} &= -\sum_{k=1}^{t-1} \frac{\partial \pi_{tk}(\theta)}{\partial \theta_i} x_{t-k}, \\
\frac{\partial^2 e_t(\theta)}{\partial \theta_i \partial \theta_j} &= -\sum_{k=1}^{t-1} \frac{\partial^2 \pi_{tk}(\theta)}{\partial \theta_i \partial \theta_j} x_{t-k}, \\
\frac{\partial^3 e_t(\theta)}{\partial \theta_i \partial \theta_j \partial \theta_l} &= -\sum_{k=1}^{t-1} \frac{\partial^3 \pi_{tk}(\theta)}{\partial \theta_i \partial \theta_j \partial \theta_l} x_{t-k},
\end{aligned} \tag{3.4}$$

for $i, j, l = 1, ..., m$.

On the other hand, for the pure moving average representation we have

$$x_t = e_t(\theta) + \sum_{k=1}^{t-1} \psi_{tk}(\theta) e_{t-k}(\theta), \tag{3.5}$$

where the coefficients $\psi_{tk}(\theta) = \psi_{tk}^{(k)}(\theta)$, $k = 0, 1, ..., t-1$, can be obtained from the autoregressive and moving average coefficients by using the following recurrences (see Mélard, 1985, pp. 36-38):

$$\begin{aligned}
\psi_{t0}^{(0)}(\theta) &= I_r, \quad \psi_{tj}^{(0)}(\theta) = B_{tj}(\theta), \quad \widetilde{\psi}_{tj}^{(0)}(\theta) = A_{tj}(\theta), \quad \text{for} \quad j = 1, ..., t-1, \\
\psi_{tj}^{(k)}(\theta) &= \psi_{tj}^{(k-1)}(\theta) + \widetilde{\psi}_{tk}^{(k-1)}(\theta) B_{t-k,j-k}(\theta), \quad \text{for} \quad j = k, ..., t-1, \\
\widetilde{\psi}_{tj}^{(k)}(\theta) &= \widetilde{\psi}_{tj}^{(k-1)}(\theta) + \widetilde{\psi}_{tk}^{(k-1)}(\theta) A_{t-k,j-k}(\theta), \quad \text{for} \quad j = k+1, ..., t-1,
\end{aligned}$$

for each $k = 1, ..., t-1$. Hence for $\theta = \theta^0$ we have

$$x_t = g_t \epsilon_t + \sum_{k=1}^{t-1} \psi_{tk} g_{t-k} \epsilon_{t-k},$$

where we denote $\psi_{tk} = \psi_{tk}(\theta^0)$. Then, by using (3.5), $e_t(\theta)$ and its first three derivatives in (3.4) can be written as a pure moving average in terms of the innovations process:

$$e_t(\theta) = g_t\epsilon_t + \sum_{k=1}^{t-1} \psi_{t0k}(\theta, \theta^0) g_{t-k}\epsilon_{t-k}, \tag{3.6}$$

$$\frac{\partial e_t(\theta)}{\partial \theta_i} = \sum_{k=1}^{t-1} \psi_{tik}(\theta, \theta^0) g_{t-k}\epsilon_{t-k}, \tag{3.7}$$

$$\frac{\partial^2 e_t(\theta)}{\partial \theta_i \partial \theta_j} = \sum_{k=1}^{t-1} \psi_{tijk}(\theta, \theta^0) g_{t-k}\epsilon_{t-k}, \tag{3.8}$$

$$\frac{\partial^3 e_t(\theta)}{\partial \theta_i \partial \theta_j \partial \theta_l} = \sum_{k=1}^{t-1} \psi_{tijlk}(\theta, \theta^0) g_{t-k}\epsilon_{t-k}, \tag{3.9}$$

for $i, j, l = 1, ..., m$, where the coefficients $\psi_{t0k}(\theta, \theta^0)$, $\psi_{tik}(\theta, \theta^0)$, $\psi_{tijk}(\theta, \theta^0)$ and $\psi_{tijlk}(\theta, \theta^0)$ are obtained from the autoregressive and moving average coefficients by the following relations:

$$\begin{aligned}
\psi_{t0k}(\theta, \theta^0) &= \psi_{tk}(\theta^0) - \sum_{u=1}^{k} \pi_{tu}(\theta)\psi_{t-u,k-u}, \\
\psi_{tik}(\theta, \theta^0) &= -\sum_{u=1}^{k} \frac{\partial \pi_{tu}(\theta)}{\partial \theta_i}\psi_{t-u,k-u}, && (3.10) \\
\psi_{tijk}(\theta, \theta^0) &= -\sum_{u=1}^{k} \frac{\partial^2 \pi_{tu}(\theta)}{\partial \theta_i \partial \theta_j}\psi_{t-u,k-u}, && (3.11) \\
\psi_{tijlk}(\theta, \theta^0) &= -\sum_{u=1}^{k} \frac{\partial^3 \pi_{tu}(\theta)}{\partial \theta_i \partial \theta_j \partial \theta_l}\psi_{t-u,k-u}. && (3.12)
\end{aligned}$$

We denote

$$\psi_{t0k} = \psi_{t0k}(\theta^0, \theta^0), \quad \psi_{tik} = \psi_{tik}(\theta^0, \theta^0), \quad \psi_{tijk} = \psi_{tijk}(\theta^0, \theta^0) \quad \text{and} \quad \psi_{tijlk} = \psi_{tijlk}(\theta^0, \theta^0).$$

Remark 3.1 *If the process were not started at time $t = 1$, it should be necessary to impose a causality and an invertibility condition, see for example Hallin & Ingenbleek (1983) and Hallin (1986). Note that $\psi_{t0k}(\theta^0, \theta^0) = 0$, for $k \geq 1$, and 1, for $k = 0$.*

3.3 An illustration: tdVARMA$(1,1)$

Let $\{x_t : t \in \mathbf{N}\}$ be an r-vector time series satisfying

$$x_t = A_t(\theta)x_{t-1} + e_t(\theta) + B_t(\theta)e_{t-1}(\theta).$$

It is easy to see that the tdVARMA(1,1) model is considered as a special case of the model defined in (3.5) with $p = 1$ and $q = 1$. Now following Mélard (1985) and Azrak & Mélard (2015), in this special case, the coefficients of the pure moving average representation are given by:

$$\psi_{tk}(\theta) = \Big\{ \prod_{l=0}^{k-2} A_{t-l}(\theta) \Big\} \Big\{ B_{t-k+1}(\theta) + A_{t-k+1}(\theta) \Big\}, \quad \text{for} \quad k = 1, 2, ..., t-1,$$

where a product for $l = 0$ to -1 is set to I_r. The coefficients of the pure autoregressive form are

$$\pi_{tk}(\theta) = \Big\{ (-1)^{k-1} \prod_{l=0}^{k-2} B_{t-l}(\theta) \Big\} \Big\{ A_{t-k+1}(\theta) + B_{t-k+1}(\theta) \Big\},$$

so, for $i = 1, ..., m$, their derivatives are given by

$$\frac{\partial \pi_{t1}(\theta)}{\partial \theta_i} = \frac{\partial A_t(\theta)}{\partial \theta_i} + \frac{\partial B_t(\theta)}{\partial \theta_i}$$

$$\frac{\partial \pi_{t2}(\theta)}{\partial \theta_i} = -\frac{\partial B_t(\theta)}{\partial \theta_i}\{A_{t-1}(\theta) + B_{t-1}(\theta)\} - B_t(\theta)\left\{\frac{\partial A_{t-1}(\theta)}{\partial \theta_i} + \frac{\partial B_{t-1}(\theta)}{\partial \theta_i}\right\}$$

$$\begin{aligned} \frac{\partial \pi_{t3}(\theta)}{\partial \theta_i} &= \frac{\partial B_t(\theta)}{\partial \theta_i} B_{t-1}(\theta)\big\{A_{t-2}(\theta) + B_{t-2}(\theta)\big\} \\ &+ B_t(\theta)\frac{\partial B_{t-1}(\theta)}{\partial \theta_i}\big\{A_{t-2}(\theta) + B_{t-2}(\theta)\big\} \\ &+ B_t(\theta)B_{t-1}(\theta)\left\{\frac{\partial A_{t-2}(\theta)}{\partial \theta_i} + \frac{\partial B_{t-2}(\theta)}{\partial \theta_i}\right\}, \dots \end{aligned}$$

Consequently

$$\frac{\partial \pi_{tk}(\theta)}{\partial \theta_i} = (-1)^{k-1} \sum_{l=1}^{k} \left(\prod_{h=1}^{k} \chi_{t+1-h,k,l,h,i}(\theta) \right),$$

where

$$\chi_{t,k,l,h,i}(\theta) = \begin{cases} \frac{\partial \chi_{t,k,l,h}(\theta)}{\partial \theta_i} & \text{if} \quad h = l, \\ \chi_{t,k,l,h}(\theta) & \text{if} \quad h \neq l, \end{cases}$$

and

$$\chi_{t,k,l,h}(\theta) = \begin{cases} B_t(\theta) & \text{if} \quad h < k \\ A_t(\theta) + B_t(\theta) & \text{if} \quad h = k. \end{cases}$$

Then

$$\psi_{tik}(\theta, \theta^0) = \sum_{u=1}^{k} \left\{ \sum_{l=1}^{u} \left(\prod_{h=1}^{u} \chi_{t+1-h,k,l,h,i}(\theta) \right) \right\} \left\{ \prod_{h=u+1}^{k} \widetilde{\chi}_{t+1-h,k,h}(\theta^0) \right\},$$

$$\widetilde{\chi}_{t+1-h,k,h}(\theta) = \begin{cases} A_t(\theta) & \text{if} \quad h < k, \\ A_t(\theta) + B_t(\theta) & \text{if} \quad h = k. \end{cases}$$

In the univariate case these results can be found in Azrak & Mélard (2015, Chapter 4), correcting Azrak & Mélard (2006).

3.4 QMLE of tdVARMA(p, q) models: asymptotic results

Let $\{x_t : t = 1, 2, ..., n\}$ be a partial realization of length n of the process $\{x_t : t \in \mathbf{N}\}$ defined in (3.1). In the present section, we shall apply the general results of Section 2.2 to the tdVARMA(p, q) setting after formulating the minimal requirements for satisfying Assumptions $\mathbf{H}_{2.1} - \mathbf{H}_{2.4}$ (resp., $\mathbf{H}_{2.1} - \mathbf{H}_{2.5}$). The notations $Q_n(\theta), \alpha_t(\theta)$ as well as the QMLE solution (2.3) remain of course the same here.

Theorem 3.1 below establishes the strong consistency of the QMLE and further the asymptotic normality of this estimator. For convenience, we suppose that the parameters in $A_{ti}(\theta)$ for $i = 1, ..., p$, in $B_{tj}(\theta)$ for $j = 1, ..., q$, and in $g_t(\theta)$ are functionally independent. Without loss of generality we suppose that the vector θ is composed of three sub-vectors A, B and g, more concretely $\theta = (A^T, B^T, g^T)^T$, A being the sub-vector of the parameters included in $A_{ti}(\theta)$ for $i = 1, ..., p$, with dimension s_1, B the sub-vector of the parameters included in $B_{tj}(\theta)$ for $j = 1, ..., q$, with dimension s_2 and g the sub-vector of the parameters included in $g_t(\theta)$ with dimension $m - s_1 - s_2$. Let us define the following Schur or Frobenius matrix norm.

Definition 3.1 *The Schur or Frobenius norm is a matrix norm of an $m \times n$ matrix A defined as*

$$\|A\|_F = \sqrt{\operatorname{tr}(A^T A)}.$$

For further information about this matrix norm, see Golub & Van Loan (1996, p. 55).

We now introduce a set of assumptions that will allow us, as mentioned above, to use results from Section 2.2. We assume for all $t \in \mathbf{N}$:

H$_{3.1}$: The matrices $A_{ti}(\theta)$, $B_{tj}(\theta)$ and $g_t(\theta)$ are three times continuously differentiable with respect to θ, in an open set Θ which contains the true value θ^0 of θ.

H$_{3.2}$: There exist positive constants N_1, N_2, N_3, N_4, N_5 and $0 < \Phi < 1$ such that, for $\nu = 1, ..., t-1$,

$$\sum_{k=\nu}^{t-1} \|\psi_{tik}\|_F^2 < N_1 \Phi^{\nu-1}, \quad \sum_{k=\nu}^{t-1} \|\psi_{tik}\|_F^4 < N_2 \Phi^{\nu-1},$$

$$\sum_{k=\nu}^{t-1} \|\psi_{tijk}\|_F^2 < N_3 \Phi^{\nu-1}, \quad \sum_{k=\nu}^{t-1} \|\psi_{tijk}\|_F^4 < N_4 \Phi^{\nu-1},$$

$$\sum_{k=1}^{t-1} \|\psi_{tijlk}\|_F^2 < N_5, \quad i, j, l = 1, ..., m,$$

H$_{3.3}$: There exist positive constants K_1, K_2, K_3, K_4, K_5 such that

$$\left\| \left\{ \frac{\partial \Sigma_t(\theta)}{\partial \theta_i} \right\}_{\theta=\theta^0} \right\|_F^2 \leq K_1, \quad \left\| \left\{ \frac{\partial^2 \Sigma_t(\theta)}{\partial \theta_i \partial \theta_j} \right\}_{\theta=\theta^0} \right\|_F^2 \leq K_2, \quad \left\| \left\{ \frac{\partial^3 \Sigma_t^{-1}(\theta)}{\partial \theta_i \partial \theta_j \theta_l} \right\}_{\theta=\theta^0} \right\|_F^2 \leq K_3,$$

$$\left\| \left\{ \frac{\partial \Sigma_t^{-1}(\theta)}{\partial \theta_i} \right\}_{\theta=\theta^0} \right\|_F^2 \leq K_4, \quad \left\| \left\{ \frac{\partial^2 \Sigma_t^{-1}(\theta)}{\partial \theta_i \partial \theta_j} \right\}_{\theta=\theta^0} \right\|_F^2 \leq K_5, \quad i, j, l = 1, ..., m.$$

H$_{3.4}$: There exist positive constants M_1, M_2, and M_3 such that

$$E\left[({\epsilon_t}^T \epsilon_t)^4 \right] \leq M_1, \quad \left\| E\left(\epsilon_t \epsilon_t^T \otimes \epsilon_t^T \right) \right\|_F \leq M_2,$$

$$\|\kappa_t\|_F + \left\| \mathrm{vec}(\Sigma). \mathrm{vec}(\Sigma)^T \right\|_F + \|\Sigma \otimes \Sigma\|_F + \|K_{r,r}(\Sigma \otimes \Sigma)\|_F \leq M_3.$$

H$_{3.5}$: There exist positive constants m_1 and m_2 such that

$$\|g_t\|_F^2 \leq m_1, \quad \left\| \Sigma_t^{-1} \right\|_F^2 \leq m_2.$$

Furthermore, we suppose that

H$_{3.6}$:

$$\lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} \left\{ E_{\theta^0} \left(\frac{\partial e_t^T(\theta)}{\partial \theta_i} \Sigma_t^{-1}(\theta) \frac{\partial e_t(\theta)}{\partial \theta_j} \right) + \mathrm{tr}\, \frac{1}{2} \left[\Sigma_t^{-1} \frac{\partial \Sigma_t(\theta)}{\partial \theta_i} \Sigma_t^{-1} \frac{\partial \Sigma_t(\theta)}{\partial \theta_j} \right]_{\theta=\theta^0} \right\} = V_{ij},$$

for $i, j = 1, ..., m$, where the matrix $V = (V_{i,j})_{1 \le i,j \le m}$ is a strictly positive definite matrix;

$\mathbf{H}_{3.7}$:

$$\frac{1}{n^2} \sum_{d=1}^{n-1} \sum_{t=1}^{n-d} \sum_{k=1}^{t-1} \|g_{t-k}\|_F^2 \|\psi_{tik}\|_F \left\|\psi_{t+d,i,k+d}\right\|_F = O\left(\frac{1}{n}\right),$$

$$\begin{aligned}
&\frac{1}{n^2} \sum_{d=1}^{n-1} \sum_{t=1}^{n-d} \left[\sum_{k=1}^{t-1} M_{t0kk}^{jiT} \Xi_{t-k} M_{tdkk}^{ij} \right. \\
&+ \sum_{k_1=1}^{t-1} \sum_{k_2=1}^{t-1} M_{t0k_2k_1}^{jiT} K_{r,r}(\Sigma \otimes \Sigma) M_{tdk_1k_2}^{ij} \\
&\left. + \sum_{k_1=1}^{t-1} \sum_{k_2=1}^{t-1} M_{t0k_2k_1}^{jiT} (\Sigma \otimes \Sigma) M_{tdk_2k_1}^{ij} \right] = O\left(\frac{1}{n}\right),
\end{aligned}$$

with

$$\Xi_t(\Sigma) = \kappa_t(\Sigma) - \mathrm{vec}(\Sigma).\, \mathrm{vec}(\Sigma)^T - (\Sigma \otimes \Sigma) - K_{r,r}(\Sigma \otimes \Sigma),$$

and, for $k', k'' = k, k_1, k_2$,

$$M_{tfk'k''}^{ij} = \mathrm{vec}(g_{t-k'}^T \psi_{t+f,i,k'+f}^T \Sigma_{t+f}^{-1} \psi_{t+f,j,k''+f} g_{t-k''}), \quad f = 0, d.$$

Remark 3.2 *These assumptions are a generalization of those in Azrak & Mélard (2006). Note however that their assumption about upper bounds of the 4th order moment of the process is not needed in their proof. It will also not be used here, hence it is left out.*

With these assumptions in hand, we are able to show that the conditions for Theorems 2.1- 2.2 hold (see Appendix A.1 for a sketch of the proof) and hence obtain the following result.

Theorem 3.1 *Suppose that Assumptions* $\mathbf{H}_{3.1}$-$\mathbf{H}_{3.7}$ *hold. Then there exists a sequence of estimators* $\widehat{\theta}_n = (\widehat{\theta}_1, ..., \widehat{\theta}_m)^T$ *such that*

- $\widehat{\theta}_n \to \theta^0$ *a.s., and for every* $\epsilon > 0$ *there exists an event* E *with* $P_{\theta^0}(E) > 1 - \epsilon$ *and an* n_0 *such that, for* $n > n_0$ *on* E, $Q_n(\theta)$ *reaches a relative minimum at the point* $\widehat{\theta}_n$*;*

- $n^{1/2}(\widehat{\theta}_n - \theta^0) \xrightarrow{L} \mathcal{N}(0, V^{-1}WV^{-1})$, *with*

$$W = \lim_{n\to\infty} \frac{1}{4n} \sum_{t=1}^{n} E_{\theta^0} \left(\frac{\partial \alpha_t(\theta)}{\partial \theta} \frac{\partial \alpha_t(\theta)}{\partial \theta^T} \right).$$

Remark 3.3 *For the sake of simplicity in the proof of Theorem 3.1 (see Appendix A.3), we have made assumptions on* Σ_t *in addition to those on* g_t*. The proof is somewhat similar to that of Azrak & Mélard (2006) but is extended to multivariate processes. Note however several corrections with respect to that paper (e.g. the treatment of the third term of* (A.4)*;* θ^0 *and* θ *were sometimes not distinguished where they should, especially in Section 3.2) and improvements (the treatment of the last three terms of* (A.4) *is more detailed; also the role of the assumptions is better enlightened).*

4 Some examples

The two examples will show that the theory can be applied and that the assumptions can be verified.

4.1 Example 1: tdVAR(1) a generalization of Kwoun & Yajima (1986)

In the following example we discuss a generalization of Kwoun & Yajima (1986). To achieve this we consider the bivariate tdVAR(1) model

$$\begin{pmatrix} x_{t1} \\ x_{t2} \end{pmatrix} = \begin{pmatrix} A_t^{11} & A_t^{12} \\ A_t^{21} & A_t^{22} \end{pmatrix} \begin{pmatrix} x_{t-1,1} \\ x_{t-1,2} \end{pmatrix} + \begin{pmatrix} \epsilon_{t1} \\ \epsilon_{t2} \end{pmatrix}, \tag{4.1}$$

where the coefficients $A_t^{ij}(\theta)$ for $i, j = 1, 2$ are defined as

$$A_t^{ij}(\theta) = A'_{ij} \sin(\alpha_{ij} t + A''_{ij}). \tag{4.2}$$

The unknown parameters A'_{ij} and A''_{ij} are such that $A'_{ij} \in [\delta, 1-\delta]$ and $A''_{ij} \in [0, 2\pi - \delta]$ for some fixed $1/2 > \delta > 0$ and α_{ij} are known constants. Then

$$\theta = (A'_{11}, A'_{21}, A'_{12}, A'_{22}, A''_{11}, A''_{21}, A''_{12}, A''_{22})^T.$$

The numerical example proposed by Kwoun & Yajima (1986) for $r = 1$ contains a process with periodic coefficients of period 4 because $\alpha_{ij} = \pi/2$. However, it is well known, see e.g. Tiao & Grupe (1980) and Azrak & Mélard (2006), that an r-dimensional autoregressive process with periodic coefficients of period $s \in \mathbf{N}$ can be embedded into an s-dimensional stationary autoregressive process. To avoid this simplification we consider coefficients $A_t^{ij}(\theta)$ either with distinct irrational periods or at least with large relatively prime periods. We check the assumptions of Theorem 3.1 in Appendix B.1 in the simplified case where we have

$$A_t(\theta) \quad = \quad \begin{pmatrix} A'_{11}\sin(at) & \frac{1}{2} \\ 0 & A'_{22}\sin(bt) \end{pmatrix}, \tag{4.3}$$

with $\theta = (A'_{11}, A'_{22})^T$. For the simulation study, we shall rather have recourse to the model

$$A_t(\theta) \quad = \quad \begin{pmatrix} A'_{11}\sin(at) & A'_{12} \\ 0 & A'_{22}\sin(bt) \end{pmatrix}, \tag{4.4}$$

with

$$\theta = (A'_{11}, A'_{12}, A'_{22})^T, \quad a = \frac{2\pi}{\sqrt{2499}} \quad \text{and} \quad b = \frac{2\pi}{\sqrt{2399}}. \tag{4.5}$$

We impose that $A'_{11}, A'_{22} \in [0, 1[$, partly similarly to Kwoun & Yajima (1986), and write $\theta^0 = (A'^0_{11}, 0.5, A'^0_{22})^T$ for the true value of θ. For simplicity, we take here $\Sigma = I_2$.

4.2 Example 2: tdVAR(1) with heteroscedasticity

Let us re-consider the model defined in (4.1)-(4.3), except that A'_{12} is no longer a parameter, with the added difficulty that the innovations are now $g_t\epsilon_t$ instead of ϵ_t. Therefore we introduce a matrix $g_t(\theta)$ and we have a bounded time-dependent covariance matrix $\Sigma_t(\theta) = g_t(\theta)\Sigma g_t^T(\theta)$. We have extended the Kwoun & Yajima (1986) parametrization by taking

$$g_t(\theta) \quad := \quad \begin{pmatrix} \exp\left(-\eta_{11}\sin(ct)\right) & 1 \\ -1 & \exp\left(-\eta_{22}\sin(ct)\right) \end{pmatrix} \tag{4.6}$$

with $c \in \mathbb{R}$. Also we use a matrix Σ which is no longer the identity matrix:

$$\Sigma := \begin{pmatrix} s_{11} & s_{12} \\ s_{12} & s_{22} \end{pmatrix}.$$

A close examination of $\Sigma_t(\theta)$ shows that if $g_t(\theta)$ were diagonal, then the correlation between the residuals would be constant, which is not very realistic for a time-dependent process. This is why we have put off-diagonal elements different from 0 in (4.6). Here, the vector θ reduces to

$$\theta = (A'_{11}, A'_{22}, \eta_{11}, \eta_{22})^T.$$

The assumptions of Theorem 3.1 are checked in Appendix B.2, and simulation results shown in Section 5.2.

5 Simulation results

5.1 Example 1: tdVAR(1) a generalization of Kwoun & Yajima (1986)

The simulation experiment is performed in Matlab by using the program, which we call AJM, described in Alj et al. (2015c) and based on a special case of tdVAR(1) process defined in (4.4)-(4.5), with $A'_{11} = 0.8$ and $A'_{22} = -0.9$ and $(\epsilon_{t1}, \epsilon_{t2})^T$ has a bivariate normal distribution with covariance matrix $\Sigma = I_2$. A simulated series using these specifications is shown in Fig. 1.

The true value of θ is

$$\theta^0 = (A'^0_{11}, A^0_{12}, A'^0_{22})^T = (0.8, 0.5, -0.9)^T,$$

We take

$$\theta^i = (0.1 \quad 0.1 \quad 0.1)^T.$$

as initial value of θ.

The experiment was replicated 1000 times. The results are summarized in Table 1. As the sample size becomes larger, we can see that

- the averages of the estimates become closer to their true value in accordance

Table 1: Estimation results for the model (4.1) under (4.4)-(4.5) via the program AMJ, where lines (a) give the averages of the parameter estimates, lines (b) give the averages across simulations of estimated standard errors of the corresponding estimates for the 1000 replications, lines (c) the sample standard deviations of the corresponding estimates for the 1000 replications and lines (d) give percentages of simulations where we reject the hypothesis $H_0(\theta_i = \theta_i^0)$ at significance level 5%.

Sample size		$\widehat{A'}_{11}$	$\widehat{A'}_{12}$	$\widehat{A'}_{22}$	$\widehat{\Sigma}_{11}$	$\widehat{\Sigma}_{12}$	$\widehat{\Sigma}_{22}$
		$A'^0_{11} = 0.8$	$A'^0_{12} = 0.5$	$A'^0_{22} = -0.9$	$\Sigma_{11} = 1$	$\Sigma_{12} = 0$	$\Sigma_{22} = 1$
25	(a)	0.7481	0.5014	-0.8036	0.9804	-0.0026	1.0760
	(b)	0.2023	0.1543	0.2118	-	-	-
	(c)	0.2161	0.1654	0.2226	-	-	-
	(d)	7.1	7.7	4.8	-	-	-
50	(a)	0.7714	0.5035	-0.8410	0.9845	0.0022	1.0392
	(b)	0.1397	0.1049	0.1355	-	-	-
	(c)	0.1344	0.1112	0.1525	-	-	-
	(d)	4.5	6	6.6	-	-	-
100	(a)	0.7855	0.4975	-0.8650	0.9927	0.0065	1.0224
	(b)	0.0963	0.0735	0.0926	-	-	-
	(c)	0.0995	0.0733	0.0964	-	-	-
	(d)	5.4	4.8	5	-	-	-
200	(a)	0.7905	0.4984	-0.8905	0.9959	0.0037	1.0087
	(b)	0.0677	0.0510	0.0628	-	-	-
	(c)	0.0713	0.0513	0.0635	-	-	-
	(d)	5.4	5.7	4.2	-	-	-
400	(a)	0.7976	0.5000	-0.8932	0.9946	-0.0001	1.0073
	(b)	0.0474	0.0358	0.0440	-	-	-
	(c)	0.0471	0.0363	0.0448	-	-	-
	(d)	5.2	4.7	5.1	-	-	-

with the theory,

- the sample standard deviation on line (b) becomes also closer to the averages across simulations of estimated standard errors in line (c) showing that the standard errors are well estimated, and
- the percentage of rejecting the hypothesis H_0 is close to 5%.

We compare, for the sample size $n = 400$, a histogram of the 1000 replications of $\widehat{\theta}_i$ to the normal probability curve with mean equal to θ^0 and standard deviation given

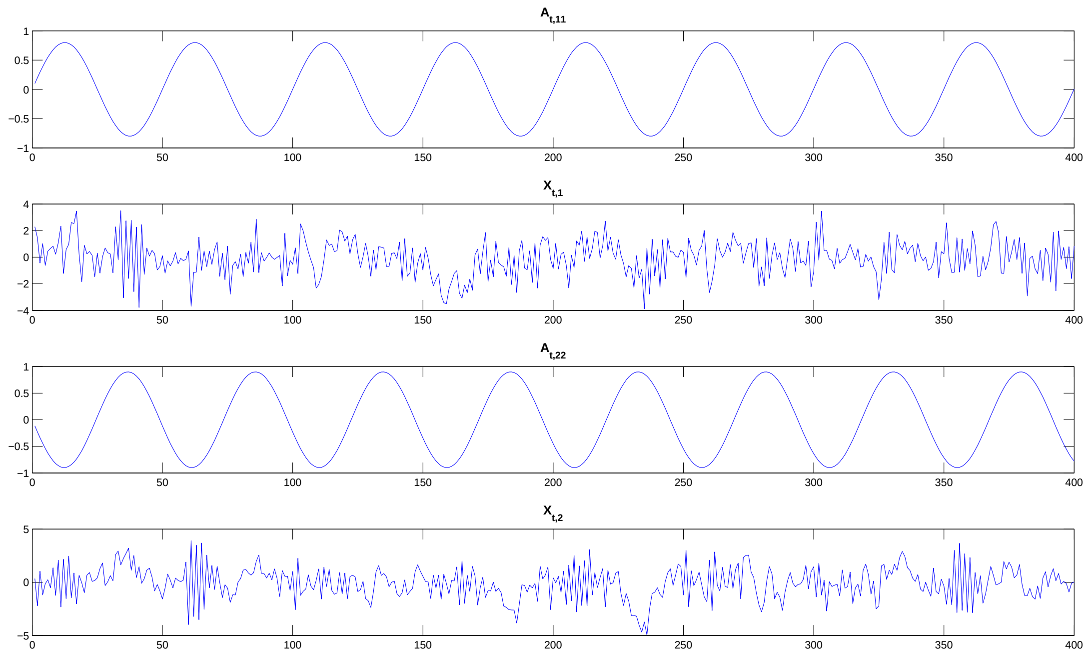

Figure 1: Time plots of the coefficients and the simulated tdVAR(1) generated by the process defined in (4.4) of length $n = 400$.

in line (b) of Table 1. As we can see from the corresponding Figure 2, this histogram shows empirically consistency and normality of the estimates.

5.2 Example 2: tdVAR(1) with heteroscedasticity

We keep the bivariate model defined by (4.1)-(4.3), with the same numerical values for a and b but without $A'_{12}(\theta)$, with $g_t\epsilon_t$ instead of ϵ_t and a covariance matrix $\Sigma_t(\theta) = g_t(\theta)\Sigma g_t^T(\theta)$ bounded but time-dependent, where

$$g_t(\theta) := \begin{pmatrix} \exp\left(-\eta_{11}\sin(ct)\right) & 1 \\ -1 & \exp\left(-\eta_{22}\sin(ct)\right) \end{pmatrix}, \tag{5.1}$$

with $\Sigma_{11} = \Sigma_{22} = 1$, $\Sigma_{12} = 0.5$, $c = 2\pi/25$ and $\eta_{11} = 1$, $\eta_{22} = -1$ so that the innovation correlation coefficient varies between -0.8 and 0.8.

Again, the number of replications is 1000. The results are presented in Table 2. Moreover a program for computing the asymptotic information matrix on the basis of the formulae given in Appendix B.2 gave, for $n = 50$ for example, the standard errors

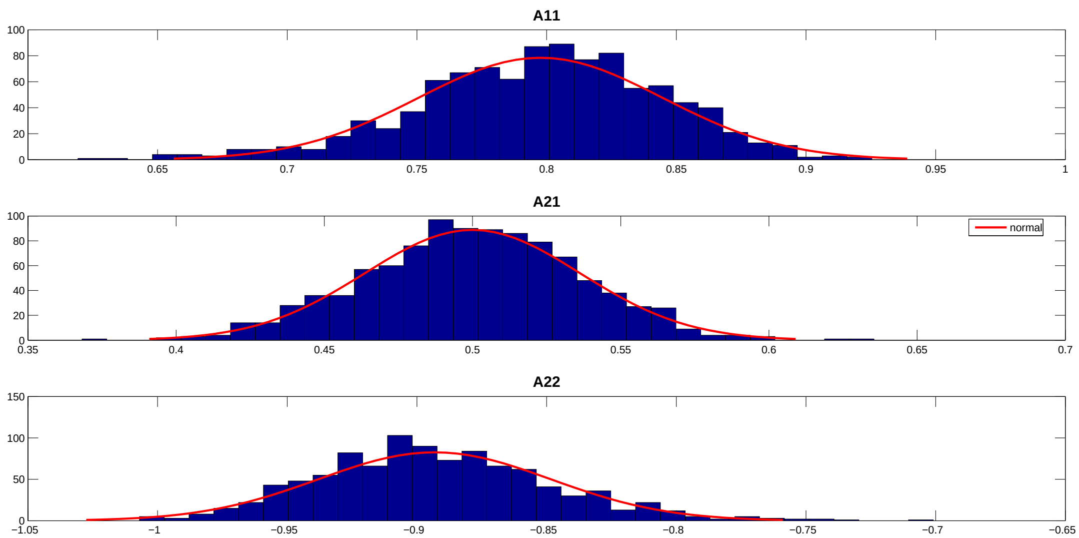

Figure 2: Histograms with the normal density function of 1000 replications of the parameters with $n = 400$.

Table 2: Estimation results for the model (4.1) under (4.4)-(4.5) and (5.1), via the program AMJ, where columns (a) give the averages of the parameter estimates and columns (d) give percentages of 1000 simulations where we reject the hypothesis $H_0(\theta_i = \theta_i^0)$ at significance level 5%.

Sample size	$\widehat{A'}_{11}$ ($A_{11}^0 = 0.8$)		$\widehat{A'}_{22}$ ($A_{22}^0 = -0.9$)		$\widehat{\eta}_{11}$ ($\eta_{11}^0 = 1.0$)		$\widehat{\eta}_{22}$ ($\eta_{22}^0 = -1.0$)	
	(a)	(d)	(a)	(d)	(a)	(d)	(a)	(d)
25	0.7671	3.8	−0.8567	4.5	0.9897	3.2	−0.9848	5.1
50	0.7808	4.2	−0.8766	5.0	0.9913	2.9	−0.9964	5.8
100	0.7910	4.4	−0.8864	5.8	0.9977	4.1	−0.9975	6.7
200	0.7963	5.3	−0.8931	5.1	0.9997	5.3	−1.0000	6.4
400	0.7972	6.3	−0.8970	4.2	0.9980	4.7	−0.9984	5.6

0.0905, 0.908, 0.1995, 0.1995, whereas the averages of the standard errors estimated by the QMLE program were 0.0963, 0.1217, 0.2027, 0.1516, respectively, and the standard deviations of the 1000 estimates were 0.0917, 0.1227, 0.1879, 0.1587.

ACKNOWLEDGMENTS

Christophe Ley, who is also a member of ECARES, thanks the Fonds National de la

Recherche Scientifique, Communauté Française de Belgique, for financial support via a Mandat de Chargé de Recherche FNRS. Guy Mélard has benefited from a Belgian research grant F.R.S.-FNRS 1.5.261.09.

References

Alj, A., Azrak, R. & Mélard, G. (2015a). Technical appendix to "Asymptotic properties of QML estimators for VARMA models with time-dependent coefficients, Parts 1 and 2". Technical report.

Alj, A., Azrak, R., & Mélard, G. (2015b). Asymptotic properties of QML estimators for VARMA models with time-dependent coefficients, part 2. In *Time series analysis by time dependent models* (eds R. Azrak & G. Mélard). In preparation.

Alj, A., Jónasson, K. & Mélard, G. (2015c). The exact Gaussian likelihood estimation of time-dependent VARMA models. *Comput. Statist. Data Anal.*, in press.

Azrak, R. & Mélard, G. (2006). Asymptotic properties of quasi-likelihood estimators for ARMA models with time-dependent coefficients. *Stat. Inference Stoch. Process.* **9**, 279–330.

Azrak, R. & Mélard, G. (2011). Autoregressive models with time-dependent coefficients - A comparison with Dahlhaus approach. ECARES working paper, Université Libre de Bruxelles.

Azrak, R. & Mélard, G. (2015). *Time series analysis by time dependent models.* In preparation.

Basawa, I. V. & Lund, R. L. S. (2001). Large sample properties of parameter estimates for periodic ARMA models. *J. Time Series Anal.* **22**, 651–663.

Basawa, I. V. & Prakasa Rao, B. L. S. (1980). *Statistical inference for stochastic processes.* Academic Press, New York.

Bibi, A. & Francq, C. (2003). Consistent and asymptotically normal estimators for cyclically time-dependent linear models. *Ann. Inst. Statist. Math.* **55**, 41–68.

Brockwell, P. J. & Davis, R. A. (1991). *Time series: theory and methods.* Springer, New York.

Boubacar Mainassara, Y. & Francq, C. (2011). Estimating structural VARMA models with uncorrelated but non-independent error terms. *J. Multivariate Anal.* **102**, 496–505.

Box, G. E. P., Jenkins, G. M. & Reinsel G. C. (2008). *Time series analysis, forecasting and control*, 4th edn. Wiley, New York.

Creal, D. D., Koopman, S. J. & Lucas, A. (2013). Generalized autoregressive score models with Applications, *J. Appl. Econom.* **28**, 777-795.

Dahlhaus, R. (1996a). Maximum likelihood estimation and model selection for locally stationary processes. *J. Nonparametr. Stat.* **6**, 171–191.

Dahlhaus, R. (1996b). On the Kullback-Leibler information divergence of locally stationary processes. *Stoch. Process. Appl.* **62**, 139–168.

Dahlhaus, R. (1996c). Asymptotic statistical inference for nonstationary processes with evolutionary spectra, In *Athens Conference on applied probability and time series analysis 2* (eds P. M. Robinson & M. Rosenblatt), 145–159. Springer, New York.

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.* **25**, 1–37.

Dahlhaus, R. (2000). A likelihood approximation for locally stationary processes. *Ann. Statist.* **28**, 1762–1794.

Freiman, G. & Halberstam, H. (1988). On product of sines. *Acta Mathematica* **49**, 377–385.

Francq, C. & Raïssi, H. (2007). Multivariate portmanteau test for autoregressive models with uncorrelated but nonindependent errors. *J. Time Series Anal.* **28**, 454–470.

Grillenzoni, C. (1990). Modeling time-varying dynamical systems. *J. Amer. Statist. Assoc.* **85**, 499–507.

Golub, G. & Van Loan, C. (1996). *Matrix computations*, 3rd edn. Johns Hopkins University Press, Baltimore.

Hall, P. & Heyde, C. C. (1980). *Martingale limit theory and its application.* Academic Press, New York.

Hallin, M. (1986). Non-stationary q-dependent processes and time-varying moving average models: invertibility properties and the forecasting problem. *Adv. Appl. Probab.* **18**, 170–210.

Hallin, M. & Ingenbleek, J. F. (1983). Nonstationary Yule-Walker equations. *Statist. Probab. Lett.* **1**, 189–195.

Hamdoune, S. (1995). Etude des problèmes d'estimation de certains modèles ARMA évolutifs. Thesis presented at Université Henri Poincaré, Nancy 1.

Hannan E. J. & Deistler, M. (1988). *The statistical theory of linear systems.* Wiley, New York.

Hindrayanto, I., Koopman, S. J. & Ooms, M. (2010). Exact maximum likelihood estimation for non-stationary periodic time series models. *Comput. Statist. Data Anal.* **54**, 2641–2654.

Janous, W. & King, J. (2000). More on a sine product formula. *The Mathematical Gazette* **84**, 113–115.

Klimko, L. A. & Nelson, P. I. (1978). On conditional least squares estimation for stochastic processes. *Ann. Statist.* **6**, 629–642.

Kohn, R. (1978). Asymptotic properties of time domain Gaussian estimators. *Adv. Appl. Probab.* **10**, 339–359.

Kollo, T. & von Rosen, D. (2005). *Advanced multivariate statistics with matrices.* Springer Verlag, New York.

Kwoun, G. H. & Yajima, Y. (1986). On an autoregressive model with time-dependent coefficients. *Ann. Inst. Statist. Math. Part A* **38**, 297–309.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis.* Springer-Verlag, New York.

Mélard, G. (1985). *Analyse de données chronologiques.* Coll. Séminaire de mathématiques supérieures - Séminaire scientifique OTAN (NATO Advanced Study Institute) **89**, Presses de l'Université de Montréal, Montréal.

Priestley, M. B. (1988). *Non-linear and non-stationary time series analysis.* Academic Press, New York.

Quenouille, M. H. (1957). *The analysis of multiple time series*. Griffin, London.

Singh, N. & Peiris, M. S. (1987). A note on the properties of some nonstationary ARMA processes. *Stoch. Process. Appl.* **24**, 151–155.

Stout, W. F. (1974). *Almost sure convergence*. Academic Press, New York.

Subba Rao, T. (1970). The fitting of non-stationary time-series models with time dependent parameters. *J. Roy. Statist. Soc. Ser. B* **32**, 312–322.

Taniguchi, M. & Kakizawa, Y. (2000). *Asymptotic theory of statistical inference for time series*. Springer Verlag, New York.

Teräsvirta, T. & Yang, Y. (2014) Linearity and misspecification tests for vector smooth transition regression models, CREATES Research Papers; No. 2014-04, Aarhus Universitet.

Teräsvirta, T., Tjøstheim, D. & Granger C. W. J. (2010) *Modelling nonlinear economic time series*. Oxford: Oxford University Press.

Tiao, G. C. & Grupe, M. R. (1980). Hidden periodic autoregressive-moving average models in time series data. *J. Roy. Statist. Soc. Ser. B* **67**, 365–373.

Tjøstheim, D. (1986). Estimation in nonlinear time series models. *Stoch. Process. Appl.* **21**, 251–273.

Triantafyllopoulos, K. & Nason, G. P. (2007). A Bayesian analysis of moving average processes with time-varying parameters. *Comput. Statist. Data Anal.* **52**, 1025–1046.

Van Bellegem, S. & Dahlhaus, R. (2006). Semiparametric estimation by model selection for locally stationary processes. *J. Roy. Statist. Soc. Ser. B* **68**, 721–746.

Van Bellegem, S. & von Sachs, R. (2004). Forecasting economic time series with unconditional time-varying variance. *International Journal of Forecasting* **20**, 611–627.

Whittle, P. (1965). Recursive relations for predictors of non-stationary processes. *J. Roy. Statist. Soc. Ser. B* **27**, 523–532.

Appendices

A Proof of theorems

Since we proceed as in Azrak & Mélard (2006), we just sketch the proof, at least in Section A.1. Lemmas preceeded by TA refer to the Technical Appendix, Alj et al. (2015a).

A.1 Proof of Theorem 2.1 and of Theorem 2.2

As mentioned in Section 2.2, we check the following four assumptions of Klimko & Nelson (1978) for $Q_n(\theta)$ defined by (2.2).

Assumption $\mathbf{H}_{1.1}$ $n^{-1}\left\{\frac{\partial Q_n(\theta)}{\partial\theta_i}\right\}_{\theta=\theta^0} \xrightarrow{\text{a.s.}} 0$, for $i = 1, ..., m$;

Assumption $\mathbf{H}_{1.2}$ $n^{-1}\left\{\frac{\partial^2 Q_n(\theta)}{\partial\theta_i\theta_j}\right\}_{\theta=\theta^0} \xrightarrow{\text{a.s.}} V_{ij}$, for $i, j = 1, ..., m$, where $V = (V_{ij})_{1\leq i,j\leq m}$ is a strictly positive definite matrix of constants;

Assumption $\mathbf{H}_{1.3}$ $\lim_{n\to\infty}\sup_{\Delta\downarrow 0}(n\Delta)^{-1}\left|\left\{\frac{\partial^2 Q_n(\theta)}{\partial\theta_i\theta_j}\right\}_{\theta=\theta^*} - \left\{\frac{\partial^2 Q_n(\theta)}{\partial\theta_i\theta_j}\right\}_{\theta=\theta^0}\right| < \infty$ a.s., for $i, j = 1, ..., m$, where θ^* is a point of the straight line joining θ^0 to θ, such that $\|\theta - \theta^0\| < \Delta$, $0 < \Delta$.

Assumption $\mathbf{H}_{1.4}$ $n^{-1/2}\left\{\frac{\partial Q_n(\theta)}{\partial\theta}\right\}_{\theta=\theta^0} \xrightarrow{L} \mathcal{N}(0, W)$, where $W = (W_{ij})_{1\leq i,j\leq m}$ is a strictly positive definite matrix.

Remark A.1 *Assumption* $\mathbf{H}_{1.3}$ *coincides with* $\mathbf{H}_{2.4}$.

A.1.1 Proof of $\mathbf{H}_{1.1}$

By TA Lemma 4.4, we have that $\{\partial\alpha_t(\theta)/\partial\theta_i, F_t\}$ is a martingale difference sequence. Then we can use a strong law of large numbers for martingale sequences (Stout, 1974, p. 154) since $\mathbf{H}_{2.1}$ implies the condition for it, more precisely

$$\sum_{t=1}^{\infty}\frac{E_{\theta^0}\left|\frac{\partial\alpha_t(\theta)}{\partial\theta_i}\right|^p}{t^{1+p/2}} < \infty, \tag{A.1}$$

for $p = 4$. $\square$

A.1.2 Proof of $\mathbf{H}_{1.2}$

By TA Lemma 4.5, we have $\{\partial^2 \alpha_t(\theta)/\partial\theta_i\partial\theta_j - E_\theta[\partial^2\alpha_t(\theta)/\partial\theta_i\partial\theta_j], F_t\}$ is a martingale difference sequence. Then we can again use the Stout (1974) strong law of large numbers, by adapting (A.1), since $\mathbf{H}_{2.2}$ implies the condition for it for $p = 2$. Hence

$$\frac{1}{n}\sum_{t=1}^{n}\left\{\frac{\partial^2\alpha_t(\theta)}{\partial\theta_i\partial\theta_j}\right\}_{\theta=\theta^0} - \frac{1}{n}\sum_{t=1}^{n}E_{\theta^0}\left(\frac{\partial^2\alpha_t(\theta)}{\partial\theta_i\partial\theta_j}/F_{t-1}\right)\xrightarrow{\text{a.s.}} 0, \quad i,j = 1,...,m.$$

Then one half of the a.s. limit of the first term is $n^{-1}\partial^2 Q_n(\theta)/\partial\theta_i\partial\theta_j$ for $\theta = \theta^0$ and defines V_{ij}, by $\mathbf{H}_{2.3}$. $\square$

Up to now, all the assumptions of Theorem 2.1 are verified. To prove Theorem 2.2, there remains to check $\mathbf{H}_{1.4}$.

A.1.3 Proof of $\mathbf{H}_{1.4}$

We proceed by using the Central Limit Theorem for martingale difference sequences of Basawa & Prakasa Rao (1980). Using the Cramér-Rao device with any vector λ, this requires to prove that, for $\xi_t = \lambda^T[\partial\alpha_t(\theta)/\partial\theta]_{\theta=\theta^0}$, $E|\xi_t|^4$ is bounded.
This is done again using $\mathbf{H}_{2.1}$. Then, with W defined by $\mathbf{H}_{2.5}$, asymptotic normality follows with the stated covariance matrix. $\square$

A.2 Further preliminaries

In order to prove Theorem 3.1, we need the following Lemma, due to Hamdoune (1995) (see also Azrak & Mélard 2006), and a strong law of large numbers for mixingale sequences.

Lemma A.1 *Let $\{w_t, t = 1,...,n\}$ be, for each $n \in N$, a scalar process with finite second-order moments, i.e.*
Lemma A.1-i: $E(w_t^2) < \infty$
Lemma A.1-ii: $E\left(n^{-1}\sum_{t=1}^n w_t^2\right) = O(n^{-\delta})$ with $\delta > 0$.
Then, $n^{-1}\sum_{t=1}^n w_t$ converges a.s. to zero when n tends to infinity.

We also need a strong law of large numbers for mixingale sequences, e.g. Hall & Heyde (1980, Theorem 2.21) in the special case where their sequence $b_n = n$. Let us recall

the definition from Hall & Heyde (1980, Section 2.3).

Definition A.1 *Let $\{w_t, t \geq 1\}$ be square-integrable random variables on a probability space (Ω, F, P) and $\{F_t, -\infty < t < \infty\}$ be an increasing sequence of σ-fields of F. Then $\{w_t, F_t\}$ is a L_2-mixingale sequence if for sequences of nonnegative constants ψ_ν and c_t where $\psi_\nu \to 0$ as $\nu \to \infty$, we have*
i. $E\{E(w_t|F_{t-\nu})^2\} \leq \psi_\nu c_t$, *and*
ii. $E(w_t - E(w_t|F_{t+\nu}))^2 \leq \psi_{\nu+1} c_t$.

Lemma A.2 *If $\{w_t, F_t\}$ is a L_2-mixingale sequence, and if $\sum_{t=1}^n c_t^2 < \infty$ and $\psi_n = O(n^{-1/2}(\log n)^{-2})$ as $n \to \infty$, then $n^{-1}\sum_{t=1}^n w_t \xrightarrow{a.s.} 0$ as $n \to \infty$.*

A.3 Proof of Theorem 3.1

First of all, as shown in Section 3.2, assumption $\mathbf{H}_{3.1}$ is used to define the ψ's used in $\mathbf{H}_{3.2}$ and in the derivatives in the other assumptions. The idea is to check the five assumptions of Theorems 2.1 and 2.2. Then $\mathbf{H}_{2.1}$ and $\mathbf{H}_{2.2}$ are direct consequences of TA Lemma 2.1 (using TA Lemma 4.11) and TA Lemma 2.2 (using TA Lemma 4.12), respectively. This makes use of assumptions $\mathbf{H}_{3.2}$, $\mathbf{H}_{3.3}$, $\mathbf{H}_{3.4}$ and $\mathbf{H}_{3.5}$.

A.3.1 Proof of $\mathbf{H}_{2.3}$

Let us consider the process Z_{tij} defined by

$$Z_{tij} \quad = \quad \left\{\frac{\partial e_t^T(\theta)}{\partial\theta_i}\Sigma_t^{-1}(\theta)\frac{\partial e_t(\theta)}{\partial\theta_j}\right\}_{\theta=\theta^0} - E_{\theta^0}\left(\frac{\partial e_t^T(\theta)}{\partial\theta_i}\Sigma_t^{-1}(\theta)\frac{\partial e_t(\theta)}{\partial\theta_j}\right).$$

By using TA Lemma 4.13 and assumptions $\mathbf{H}_{3.2}$, $\mathbf{H}_{3.4}$, $\mathbf{H}_{3.5}$ and $\mathbf{H}_{3.7}$, the two assumptions i and ii of Lemma A.1 are fulfilled for Z_{tij}, hence

$$\frac{1}{n}\sum_{t=1}^n\left[\left\{\frac{\partial e_t^T(\theta)}{\partial\theta_j}\Sigma_t^{-1}(\theta)\frac{\partial e_t(\theta)}{\partial\theta_j}\right\}_{\theta=\theta^0} - E_{\theta^0}\left(\frac{\partial e_t^T(\theta)}{\partial\theta_j}\Sigma_t^{-1}(\theta)\frac{\partial e_t(\theta)}{\partial\theta_j}\right)\right] \xrightarrow{\text{a.s.}} 0 \tag{A.2}$$

when $n \to \infty$. However, from TA Lemma 4.5, we have for $i, j = 1, ..., m$:

$$\frac{1}{2n}\sum_{t=1}^n E_{\theta^0}\left(\frac{\partial^2\alpha_t(\theta)}{\partial\theta_i\partial\theta_j}/F_{t-1}\right) \quad = \quad \frac{1}{n}\sum_{t=1}^n\left\{\frac{\partial e_t^T(\theta)}{\partial\theta_i}\Sigma_t^{-1}(\theta)\frac{\partial e_t(\theta)}{\partial\theta_j}\right\}_{\theta=\theta^0}$$

$$+\frac{1}{2n}\sum_{t=1}^{n}\operatorname{tr}\left\{\Sigma_t^{-1}(\theta)\frac{\partial\Sigma_t(\theta)}{\partial\theta_i}\Sigma_t^{-1}(\theta)\frac{\partial\Sigma_t(\theta)}{\partial\theta_j}\right\}_{\theta=\theta^0}. \tag{A.3}$$

Then (A.2) implies that the a.s. limit of (A.3) for $n \to \infty$ will be equal to

$$\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}E_{\theta^0}\left(\frac{\partial e_t^T(\theta)}{\partial\theta_i}\Sigma_t^{-1}(\theta)\frac{\partial e_t(\theta)}{\partial\theta_j}\right)$$
$$+\lim_{n\to\infty}\frac{1}{2n}\sum_{t=1}^{n}\operatorname{tr}\left\{\Sigma_t^{-1}(\theta)\frac{\partial\Sigma_t(\theta)}{\partial\theta_i}\Sigma_t^{-1}(\theta)\frac{\partial\Sigma_t(\theta)}{\partial\theta_j}\right\}_{\theta=\theta^0},$$

and this is V_{ij} as defined in $\mathbf{H}_{3.6}$.

A.3.2 Proof of $\mathbf{H}_{2.4}$

By using TA Lemma 2.3, it suffices to show that

$$\lim_{n\to\infty}\frac{1}{n}\left|\sum_{t=1}^{n}\left\{\frac{\partial^3\alpha_t(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_l}\right\}_{\theta=\theta^0}\right| \quad < \quad \infty \qquad \text{a.s.} \quad \text{for} \quad i,j,l = 1,...,m.$$

As a consequence of TA Lemma 2.4, the expression on the left-hand side can be bounded by

$$\tilde{\Phi}_1 + \frac{1}{n}\sum_{t=1}^{n}\tilde{\Phi}_{2t} + \frac{1}{n}\sum_{t=1}^{n}\tilde{\Psi}_{1t} + \frac{1}{n}\sum_{t=1}^{n}\tilde{\Psi}_{2t} + \frac{1}{n}\sum_{t=1}^{n}\tilde{\Psi}_{3t}, \tag{A.4}$$

where $\tilde{\Phi}_1$ is shown in TA Lemma 4.14 (using assumptions $\mathbf{H}_{3.3}$ and $\mathbf{H}_{3.5}$) to be bounded, and the last four terms $\tilde{\Phi}_{2t}$, $\tilde{\Psi}_{1t}$, $\tilde{\Psi}_{2t}$, and $\tilde{\Psi}_{3t}$ are briefly described now:

- $\tilde{\Phi}_{2t}$ contains terms like $e_t^T(\theta)\ (\partial^3\Sigma_t^{-1}(\theta)/\partial\theta_i\partial\theta_j\partial\theta_l)\ e_t(\theta)$;
- $\tilde{\Psi}_{1t}$ contains terms like $e_t^T(\theta)\ (\partial^2\Sigma_t^{-1}(\theta)/\partial\theta_i\partial\theta_j)\ (\partial e_t(\theta)/\partial\theta_l)$, $e_t^T(\theta)\ (\partial\Sigma_t^{-1}(\theta)/\partial\theta_i)$ $(\partial e_t^2(\theta)/\partial\theta_j\partial\theta_l)$, and $e_t^T(\theta)\ \Sigma_t^{-1}(\theta)\ (\partial e_t^3(\theta)/\partial\theta_i\partial\theta_j\partial\theta_l)$, where $i,j,l = 1,...,m$;
- the last two terms $\tilde{\Psi}_{2t}$ and $\tilde{\Psi}_{3t}$ contain respectively terms like $(\partial e_t^T(\theta)/\partial\theta_i)$ $(\partial\Sigma_t^{-1}(\theta)/\partial\theta_l)$ $(\partial e_t(\theta)/\partial\theta_j)$ and $(\partial^2 e_t^T(\theta)/\partial\theta_i\partial\theta_j)$ $\Sigma_t^{-1}(\theta)$ $(\partial e_t(\theta)/\partial\theta_l)$, where $i,j,l = 1,...,m$.

For $i, j, l = 1, ..., m$, let us define the following six sequences of random variables

$$\begin{aligned}
X_t^{ilj} &= \left(\frac{\partial e_t^T(\theta)}{\partial \theta_i}\frac{\partial \Sigma_t^{-1}(\theta)}{\partial \theta_l}\frac{\partial e_t(\theta)}{\partial \theta_j}\right)_{\theta=\theta^0} - E_{\theta^0}\left[\frac{\partial e_t^T(\theta)}{\partial \theta_i}\frac{\partial \Sigma_t^{-1}(\theta)}{\partial \theta_l}\frac{\partial e_t(\theta)}{\partial \theta_j}\right], \\
Y_t^{ijl} &= \left(\frac{\partial e_t^T(\theta)}{\partial \theta_i \partial \theta_j}\Sigma_t^{-1}(\theta)\frac{\partial e_t(\theta)}{\partial \theta_l}\right)_{\theta=\theta^0} - E_{\theta^0}\left[\frac{\partial e_t^T(\theta)}{\partial \theta_i \partial \theta_j}\Sigma_t^{-1}(\theta)\frac{\partial e_t(\theta)}{\partial \theta_l}\right], \\
Z_t &= \left(e_t^T(\theta)\frac{\partial^3 \Sigma_t^{-1}(\theta)}{\partial \theta_i \partial \theta_j \partial \theta_l}e_t(\theta)\right)_{\theta=\theta^0} - E_{\theta^0}\left[e_t^T(\theta)\frac{\partial^3 \Sigma_t^{-1}(\theta)}{\partial \theta_i \partial \theta_j \partial \theta_l}e_t(\theta)\right], \\
W_t^{(1)ijl} &= \left(e_t^T(\theta)\frac{\partial^2 \Sigma_t^{-1}(\theta)}{\partial \theta_i \partial \theta_j}\frac{\partial e_t(\theta)}{\partial \theta_l}\right)_{\theta=\theta^0} - E_{\theta^0}\left[e_t^T(\theta)\frac{\partial^2 \Sigma_t^{-1}(\theta)}{\partial \theta_i \partial \theta_j}\frac{\partial e_t(\theta)}{\partial \theta_l}\right],
\end{aligned}$$

and $W_t^{(2)ijl}$ and $W_t^{(3)ijl}$ defined similarly by replacing $(\partial^2 \Sigma_t^{-1}(\theta)/\partial \theta_i \partial \theta_j)(\partial e_t(\theta)/\partial \theta_l)$ with, respectively, $(\partial \Sigma_t^{-1}(\theta)/\partial \theta_i)(\partial e_t^2(\theta)/\partial \theta_j \partial \theta_l)$ and $\Sigma_t^{-1}(\theta)(\partial e_t^3(\theta)/\partial \theta_i \partial \theta_j \partial \theta_l)$.

First, we have that $E(Z_t/F_{t-1}) = 0$ and, according to TA Lemma 4.15 (using assumptions $\mathbf{H}_{3.3}$, $\mathbf{H}_{3.4}$ and $\mathbf{H}_{3.5}$) , that $E(Z_t^2)$ is uniformly bounded by a constant. Then, the strong law of large numbers (Stout, 1974, p. 154) implies that

$$\frac{1}{n}\sum_{t=1}^{n} Z_t \overset{\text{a.s.}}{\to} 0.$$

The arguments for the other sequences are more involved. From TA Lemmas 4.16, 4.18 and 4.20 we have that $\{W_t^{(q)ijl}, F_t\}$, $q = 1, 2, 3$, $\{X_t^{ilj}, F_t\}$ and $\{Y_t^{ijl}, F_t\}$ are L_2-mixingale sequences. From TA Lemmas 4.17, 4.19 and 4.21 we have that these L_2-mixingale sequences $\{W_t^{(q)ijl}, F_t\}$, $q = 1, 2, 3$, $\{X_t^{ilj}, F_t\}$ and $\{Y_t^{ijl}, F_t\}$ fulfil the conditions in Lemma A.2, the strong law of large numbers for a mixingale sequence (Hall & Heyde, 1980, p. 41, Theorem 2.21). This makes use of assumptions $\mathbf{H}_{3.2}$, $\mathbf{H}_{3.3}$, $\mathbf{H}_{3.4}$ and $\mathbf{H}_{3.5}$. Hence

$$n^{-1}\sum_{t=1}^{n} W_t^{(1)ijl} \overset{\text{a.s.}}{\to} 0, \quad n^{-1}\sum_{t=1}^{n} W_t^{(2)ijl} \overset{\text{a.s.}}{\to} 0, \quad n^{-1}\sum_{t=1}^{n} W_t^{(3)ijl} \overset{\text{a.s.}}{\to} 0,$$
$$n^{-1}\sum_{t=1}^{n} X_t^{ilj} \overset{\text{a.s.}}{\to} 0, \quad \text{and} \quad n^{-1}\sum_{t=1}^{n} Y_t^{ijl} \overset{\text{a.s.}}{\to} 0,$$

and this for $i, j, l = 1, ..., m$.

Consequently, the proof is completed so that $\mathbf{H}_{2.4}$ is checked. Then for every $\epsilon > 0$, there exists an event E with $P_{\theta^0}(E) > 1 - \epsilon$ and an n_0 such that, for $n > n_0$ on E, $Q_n(\theta)$ reaches a relative minimum at the point $\widehat{\theta}_n$. Consequently, there exists

an estimator $\hat{\theta}_n$ such that $\hat{\theta}_n \overset{\text{a.s.}}{\to} \theta^0$ as $n \to \infty$.

A.3.3 Proof of $\mathbf{H}_{2.5}$

From TA Lemma 4.10 we can determine the explicit form of the left-hand side of $\mathbf{H}_{2.5}$ for all $1 \leq i, j \leq m$:

$$\frac{4}{n}\sum_{t=1}^{n}\left[\left\{\frac{\partial e_t^T(\theta)}{\partial \theta_j}\Sigma_t^{-1}(\theta)\frac{\partial e_t(\theta)}{\partial \theta_i}\right\}_{\theta=\theta^0} - E_{\theta^0}\left(\frac{\partial e_t^T(\theta)}{\partial \theta_j}\Sigma_t^{-1}(\theta)\frac{\partial e_t(\theta)}{\partial \theta_i}\right)\right]$$
$$+\frac{2}{n}\sum_{t=1}^{n}\left[\left\{\frac{\partial e_t^T(\theta)}{\partial \theta_i}\right\}_{\theta=\theta^0} - E_{\theta^0}\left(\frac{\partial e_t^T(\theta)}{\partial \theta_i}\right)\right]K_{t,j}$$
$$+\frac{2}{n}\sum_{t=1}^{n}\left[\left\{\frac{\partial e_t^T(\theta)}{\partial \theta_j}\right\}_{\theta=\theta^0} - E_{\theta^0}\left(\frac{\partial e_t^T(\theta)}{\partial \theta_j}\right)\right]K_{t,i}, \tag{A.5}$$

where $K_{t,i}$ is defined in TA Lemma 4.10. While checking $\mathbf{H}_{2.3}$, we have shown that

$$\frac{4}{n}\sum_{t=1}^{n}\left[\left\{\frac{\partial e_t^T(\theta)}{\partial \theta_j}\Sigma_t^{-1}(\theta)\frac{\partial e_t(\theta)}{\partial \theta_i}\right\}_{\theta=\theta^0} - E_{\theta^0}\left(\frac{\partial e_t^T(\theta)}{\partial \theta_j}\Sigma_t^{-1}(\theta)\frac{\partial e_t(\theta)}{\partial \theta_i}\right)\right] \overset{\text{a.s.}}{\to} 0.$$

There remains to prove that the second and third terms of (A.5) also tend a.s. to zero. To achieve that, let us consider

$$\tilde{Z}_{t,ij}(\theta) = \left(\left\{\frac{\partial e_t^T(\theta)}{\partial \theta_j}\right\}_{\theta=\theta^0} - E_{\theta}\left(\frac{\partial e_t^T(\theta)}{\partial \theta_j}\right)\right)K_{t,i},$$

for $i, j = 1, ..., m$. Then, by TA Lemma 4.22 (involving assumptions $\mathbf{H}_{3.3}$, $\mathbf{H}_{3.4}$, $\mathbf{H}_{3.5}$ and $\mathbf{H}_{3.7}$), the two assumptions i and ii of Lemma A.1 are verified, entailing that the last two terms of (A.5) also tend to zero almost surely.

As a conclusion, the asymptotic convergence of the estimator $\hat{\theta}_n$ towards the normal distribution is ensured and the proof of Theorem 3.1 is achieved. □

B Assumptions checking

We check the assumptions of Theorem 3.1 for the two examples of Section 4, hereby providing a theoretical foundation for the simulation results of Section 5. Since several of the assumptions are somewhat similar, we have only covered once each argument. Also, since Example 2 is a generalization of Example 1, we have avoided to repeat some of the verifications when they are too similar.

B.1 Example 4.1

B.1.1 Assumption $\mathbf{H}_{3.1}$

Trivial

B.1.2 Assumption $\mathbf{H}_{3.2}$

In order check this hypothesis, we shall have recourse to the results of Section 3.2. In this example the coefficients of the pure moving average representation of (4.1) are given by

$$\psi_{tk}(\theta) = \prod_{l=0}^{k-1} A_{t-l}(\theta),$$

for $k = 1, 2, ..., t-1$. The coefficients of the pure autoregressive form are

$$\pi_{t1}(\theta) = A_t(\theta),$$

and $\pi_{tk}(\theta) = 0$ if $k = 2, ..., t-1$.

Then, by using (3.10)-(3.12), we can calculate $\psi_{tik}(\theta, \theta^0)$, $\psi_{tijk}(\theta, \theta^0)$ and $\psi_{tijlk}(\theta, \theta^0)$. For example, denoting $A_t^{(k-1)} = \prod_{l=1}^{k-1} A_{t-l}(\theta^0)$, and its (i, j) element $A_{t,i,j}^{(k-1)}$, $i, j = 1, 2$, we have

$$\psi_{tik}(\theta, \theta^0) = \frac{\partial \pi_{t1}(\theta)}{\partial \theta_i} \prod_{l=1}^{k-1} A_{t-l}(\theta^0) \quad = \quad \frac{\partial A_t(\theta)}{\partial \theta_i} A_t^{(k-1)}, \tag{B.1}$$

$$\psi_{tijk}(\theta, \theta^0) = \frac{\partial^2 \pi_{t1}(\theta)}{\partial \theta_i \partial \theta_j} \prod_{l=1}^{k-1} A_{t-l}(\theta^0) \quad = \quad \frac{\partial^2 A_t(\theta)}{\partial \theta_i \partial \theta_j} A_t^{(k-1)}, \tag{B.2}$$

$$\psi_{tijlk}(\theta, \theta^0) = \frac{\partial^3 \pi_{t1}(\theta)}{\partial \theta_i \partial \theta_j \partial \theta_l} \prod_{l=1}^{k-1} A_{t-l}(\theta^0) \quad = \quad \frac{\partial^3 A_t(\theta)}{\partial \theta_i \partial \theta_j \partial \theta_l} A_t^{(k-1)}, \tag{B.3}$$

for $k = 1, 2, ..., t-1$, where $A_t^{(0)} = I_2$. Obviously, checking the assumptions in the general setup happens to be a complicated and tedious task, hence, for the sake of simplification, we shall consider the model defined in (4.3). It can be shown by

induction that

$$A_t^{(k-1)} = \begin{pmatrix} A_{t,1,1}^{(k-1)} & A_{t,1,2}^{(k-1)} \\ 0 & A_{t,2,2}^{(k-1)} \end{pmatrix}, \tag{B.4}$$

for $k \geq 2$, where

$$A_{t,1,1}^{(k-1)} = (A_{11}^{\prime 0})^{k-1} \prod_{l=1}^{k-1} \sin(a(t-l)), \quad A_{t,2,2}^{(k-1)} = (A_{22}^{\prime 0})^{k-1} \prod_{l=1}^{k-1} \sin(b(t-l)), \tag{B.5}$$

$$A_{t,1,2}^{(k-1)} = \frac{1}{2} \sum_{l=1}^{k-1} (A_{11}^{\prime 0})^{k-l-1} (A_{22}^{\prime 0})^{l-1} \prod_{f=1}^{k-2} \sin(c_{lf}(t-f-\delta_{lf})), \tag{B.6}$$

and $c_{lf} = a$ and $\delta_{lf} = 0$, for $l + f \leq k - 1$, and $c_{lf} = b$ and $\delta_{lf} = 1$, for $l + f > k - 1$. For example, for $k = 4$

$$\begin{aligned} A_{t,1,2}^{(3)} = \frac{1}{2} \Big(& (A_{11}^{\prime 0})^2 \sin(a(t-1)) \sin(a(t-2)) + A_{11}^{\prime 0} A_{22}^{\prime 0} \sin(a(t-1)) \sin(b(t-3)) \\ & + (A_{22}^{\prime 0})^2 \sin(b(t-2)) \sin(b(t-3)) \Big) . \end{aligned}$$

For $i = 1$, i.e. $\theta_1 = A'_{11}$, and using (B.4) we have

$$\begin{aligned} \psi_{t1k} &= \begin{pmatrix} \sin(at) & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} A_{t,1,1}^{(k-1)} & A_{t,1,2}^{(k-1)} \\ 0 & - \end{pmatrix} \\ &= \sin(at) \begin{pmatrix} A_{t,1,1}^{(k-1)} & A_{t,1,2}^{(k-1)} \\ 0 & 0 \end{pmatrix} \end{aligned} \tag{B.7}$$

where a dash will always indicate an element that will not be used. By the same way for $\theta_2 = A'_{22}$

$$\begin{aligned} \psi_{t2k} &= \begin{pmatrix} 0 & 0 \\ 0 & \sin(bt) \end{pmatrix} \begin{pmatrix} - & - \\ 0 & A_{t,2,2}^{(k-1)} \end{pmatrix} \\ &= \sin(bt) \begin{pmatrix} 0 & 0 \\ 0 & A_{t,2,2}^{(k-1)} \end{pmatrix} . \end{aligned} \tag{B.8}$$

We define the constant

$$\Phi^{1/2} = \max \left\{ \left| A_{11}^{\prime 0} \right|, \left| A_{22}^{\prime 0} \right| \right\}, \tag{B.9}$$

which is such that $0 < \Phi < 1$. Now by using (B.5)-(B.9) and bounding sines by 1, we

can show that

$$A_{t,1,1}^{(k-1)} \leq \Phi^{\frac{k-1}{2}}, \tag{B.10}$$

$$A_{t,1,2}^{(k-1)} \leq (k-1)\Phi^{\frac{k}{2}-1} \tag{B.11}$$

and

$$A_{t,2,2}^{(k-1)} \leq \Phi^{\frac{k-1}{2}}. \tag{B.12}$$

Consequently, from (B.8) and (B.12),

$$\sum_{k=\nu}^{t-1} \|\psi_{t2k}\|_F^2 = \sum_{k=\nu}^{t-1} \left[\sin(bt)A_{t,2,2}^{(k-1)}\right]^2 \leq \sum_{k=\nu}^{t-1} \Phi^{k-1} \leq N_1\Phi^{\nu-1},$$

where $N_1 = 1/(1-\Phi)$. By the same way, from (B.7) and (B.10)-(B.11),

$$\begin{aligned} \sum_{k=\nu}^{t-1} \|\psi_{t1k}\|_F^2 &= \sum_{k=\nu}^{t-1} \sin^2(at)\left[\left(A_{t,1,1}^{(k-1)}\right)^2 + \left(A_{t,1,2}^{(k-1)}\right)^2\right] \\ &\leq \sum_{k=\nu}^{t-1} \Phi^{k-1} + \sum_{k=\nu}^{t-1}(k-1)^2\Phi^{k-2}, \end{aligned} \tag{B.13}$$

but this cannot be bounded by $N_1\Phi^{\nu-1}$ for some constant N_1, independently of ν, so a more subtle upper bound needs to be found. Using (B.6), an upper bound of the element $(1,2)$ is equal to

$$\frac{1}{2}\Phi^{k/2-1}\sum_{\ell=1}^{k-1}\prod_{f=1}^{k-2}\sin(c_{lf}(t-f-\delta_{lf})). \tag{B.14}$$

It should be possible to adapt some results on products of sines (e.g., Freiman & Halberstam, 1988, and Janous & King, 2000) to the general case. More precisely, let $Q_k = \max_{P\geq 2}\prod_{t=1}^{k}\sin(2\pi t/P)$, then $\lim_{k\to\infty}(Q_k)^{1/k} = 0.6098579....$ But it will be tricky given the specific form (B.14), and we will restrain the proof by assuming a sufficient (but not necessary) condition that $a = 2\pi/P_1$ and $b = 2\pi/P_2$, for some strictly positive integers P_1 and P_2. Then $\sin(a(t-f)) = 0$, for $t-f = g_1P_1$, for $g_1 \in \mathbb{Z}$, and $\sin(b(t-f-1)) = 0$, for $t-f-1 = g_2P_2$, for $g_2 \in \mathbb{Z}$. Let $\tilde{k} \stackrel{\text{def}}{=} P_1+P_2+1$. Then, for any given t, take the remainders g_1 and g_2 of the Euclidean divisions of t by

P_1 and of $t-1$ by P_2, respectively. These remainders are between 0 and, respectively, P_1-1 and P_2-1, hence smaller than $\tilde{k}-2$. Hence for any $k \geq \tilde{k}$, there is some integer f, either $t-g_1P_1$, or $t-1-g_2P_2$ between 0 and $k-2$ such that either $\sin(a(t-f))=0$ or $\sin(b(t-f-1))=0$ and finally each term of (B.14) vanishes. Hence, taking $\nu > \tilde{k}$, the second term of (B.13) vanishes for all t and the Frobenius norm can be bounded by some expression $N_1\Phi^{\nu-1}$ for some N_1. Note also that, under the conditions of integral periods, we have

$$\|\psi_{t1k}\|_F^2 = \|\psi_{t2k}\|_F^2 = 0, \quad k > \tilde{k}.$$

By using (B.1)-(B.2)-(B.3) the other inequalities of this assumption are checked.

B.1.3 Assumption $\mathbf{H}_{3.3}$

Trivial.

B.1.4 Assumption $\mathbf{H}_{3.4}$

Trivial.

B.1.5 Assumption $\mathbf{H}_{3.5}$

Trivial.

B.1.6 Assumption $\mathbf{H}_{3.6}$

The second term in $\mathbf{H}_{3.6}$ is equal to 0, so it remains to show that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \left(E_{\theta^0} \left(\frac{\partial e_t^T(\theta)}{\partial \theta_i} \Sigma_t^{-1}(\theta) \frac{\partial e_t(\theta)}{\partial \theta_j} \right) \right) = V_{ij}$$

for $i,j=1,2$, where the matrix $V=(V_{i,j})_{1\leq i,j\leq 2}$ is a strictly positive definite matrix. From (3.7) we have

$$\frac{\partial e_t(\theta)}{\partial \theta_1} \quad = \quad \sum_{k=1}^{t-1} \psi_{t1k}(\theta,\theta^0)\epsilon_{t-k},$$

entailing, since we have taken $\Sigma = I_2$,

$$\begin{aligned} E_{\theta^0}\left(\frac{\partial e_t^T(\theta)}{\partial\theta_1}\Sigma^{-1}\frac{\partial e_t(\theta)}{\partial\theta_1}\right) &= E_{\theta^0}\left[\left(\sum_{k=1}^{t-1}\psi_{t1k}(\theta,\theta^0)\epsilon_{t-k}\right)^T\Sigma^{-1}\sum_{k=1}^{t-1}\psi_{t1k}(\theta,\theta^0)\epsilon_{t-k}\right] \\ &= \sum_{k=1}^{t-1}\sin^2(at)\,\mathrm{tr}\left[\begin{pmatrix} A_{t,1,1}^{(k-1)} & A_{t,1,2}^{(k-1)} \\ 0 & 0 \end{pmatrix}^T\begin{pmatrix} A_{t,1,1}^{(k-1)} & A_{t,1,2}^{(k-1)} \\ 0 & 0 \end{pmatrix}\right] \\ &= \sin^2(at)\sum_{k=1}^{t-1}\left[\left(A_{t,1,1}^{(k-1)}\right)^2+\left(A_{t,1,2}^{(k-1)}\right)^2\right]. \qquad \text{(B.15)} \end{aligned}$$

The terms for $k > \tilde{k}$ vanish, as explained above. The remaining terms are bounded by Φ^{k-1}, also as above, and the sum is strictly positive, at least for some t's. Therefore, taking the average for $t = 1$ to n gives a finite strictly positive limit V_{11} when $n \to \infty$.

By the same method for $i, j = 2$

$$\begin{aligned} E_{\theta^0}\left(\frac{\partial e_t^T(\theta)}{\partial\theta_2}\Sigma^{-1}\frac{\partial e_t(\theta)}{\partial\theta_2}\right) &= E_{\theta^0}\left[\left(\sum_{k=1}^{t-1}\psi_{t2k}(\theta,\theta^0)\epsilon_{t-k}\right)^T\Sigma^{-1}\sum_{k=1}^{t-1}\psi_{t2k}(\theta,\theta^0)\epsilon_{t-k}\right] \\ &= \sum_{k=1}^{t-1}\sin^2(bt)\,\mathrm{tr}\left[\begin{pmatrix} 0 & 0 \\ 0 & A_{t,2,2}^{(k-1)} \end{pmatrix}^T\begin{pmatrix} 0 & 0 \\ 0 & A_{t,2,2}^{(k-1)} \end{pmatrix}\right] \\ &= \sin^2(bt)\sum_{k=1}^{t-1}\left(A_{t,2,2}^{(k-1)}\right)^2, \end{aligned}$$

with a similar consequence to provide a finite strictly positive limit V_{22} when $n \to \infty$.

Furthermore, it can easily be seen

$$V_{12} = V_{21} = 0.$$

Consequently, the matrix $V = (V_{i,j})_{1\le i,j\le 2}$ is strictly positive definite.

Remark B.1 *If we handle the model described in Section 5, but without the parameter A'_{12}, for the case where the number of observations is* 25*, we obtain the following standard errors from that theoretical formula for the two remaining parameters:* 0.2175 *and* 0.2303*, respectively, which largely agree with the averages drawn from the simulation results:* 0.2161 *and* 0.2226*, as shown in Table 1.*

B.1.7 Assumption $\mathbf{H}_{3.7}$

For the first part of this assumption, since g_t is bounded, it remains to show that

$$\frac{1}{n^2}\sum_{d=1}^{n-1}\sum_{t=1}^{n-d}\sum_{k=1}^{t-1}\|\psi_{tik}\|_F\,\|\psi_{t+d,i,k+d}\|_F = O\left(\frac{1}{n}\right).$$

For $i = 1$

$$\sum_{k=1}^{t-1}\|\psi_{t1k}\|_F\,\|\psi_{t+d,1,k+d}\|_F = |\sin(at)\sin(a(t+d))|$$
$$\times\sum_{k=1}^{t-1}\left[\left(A_{t,1,1}^{(k-1)}\right)^2+\left(A_{t,1,2}^{(k-1)}\right)^2\right]^{1/2}\left[\left(A_{t+d,1,1}^{(k+d-1)}\right)^2+\left(A_{t+d,1,2}^{(k+d-1)}\right)^2\right]^{1/2}. \tag{B.16}$$

To simplify the proof we assume again that $a = 2\pi/P_1$ and $b = 2\pi/P_2$, for some integers P_1 and P_2 and use $\tilde{k}$ as defined above. Then (B.16) becomes a sum for $k = 1$ to $\tilde{k} - d$.

The general term can be bounded by $[\Phi^{k-2}(1+(k-1)^2)\Phi^{k+d-2}(1+(k+d-1)^2)]^{1/2}$ which is of order $d\Phi^{d/2}$. Hence

$$\begin{aligned}\sum_{d=1}^{n-1}\sum_{t=1}^{n-d}(B.16) &= \sum_{t=1}^{n-1}\sum_{d=1}^{n-t}(B.16)\\ &\leq \sum_{t=1}^{n-1}\sum_{d=1}^{n-t}k_1(\Phi)\Phi^{d/2}d\\ &\leq k_1(\Phi)\frac{\Phi^{1/2}}{(1-\Phi^{1/2})^2}\sum_{t=1}^{n-1}1\\ &\leq k_2(\Phi)n,\end{aligned}$$

where $k_1(\Phi)$ and $k_2(\Phi)$ are constants and where we have used the formula $\sum_{j=1}^{\infty} jx^j = x/(1-x)^2$, provided $|x| < 1$. Dividing by n^2 thus gives $O(1/n)$, as requested.

Applying the same method for $i = 2$, with again a sum for $k = 1$ to $\tilde{k}$ but of $A_{t,2,2}^{(k-1)}A_{t+d,2,2}^{(k+d-1)}$ instead of the product of square roots of sums of squares in (B.16),

with a general term bounded this time by $\Phi^{(k-1)/2}\Phi^{(k+d-1)/2}$, we can show that

$$\begin{aligned}\sum_{d=1}^{n-1}\sum_{t=1}^{n-d}\sum_{k=1}^{t-1}\left\|\psi_{t2k}^{(n)}\right\|_F\left\|\psi_{t+d,2,k+d}^{(n)}\right\|_F &\leq \sum_{d=1}^{n-1}\sum_{t=1}^{n-1}k_3(\Phi)\Phi^{d/2}\\ &\leq k_4(\Phi)n,\end{aligned}$$

with other constants $k_3(\Phi)$ and $k_4(\Phi)$ and reach the same final conclusion as for $i=1$.

For the second part of this assumption, since $\Xi_t(\Sigma)$, $\Sigma\otimes\Sigma$ and $\operatorname{vec}(\Sigma)\operatorname{vec}(\Sigma)^T$ are finite constants, similarly to the first part we can show that the second part of assumption $\mathbf{H}_{3.7}$ is fulfilled.

B.2 Example 4.2

The assumptions $\mathbf{H}_{3.1}-\mathbf{H}_{3.5}$ are easily checked: $\mathbf{H}_{3.1}$ is trivial, $\mathbf{H}_{3.2}$ remains unchanged but requires the same conditions on integers P_1, P_2 as for the previous example, and $\mathbf{H}_{3.3}-\mathbf{H}_{3.5}$ are readily checked given boundedness of the sine function. It remains to discuss about $\mathbf{H}_{3.6}$ and $\mathbf{H}_{3.7}$ which we are going to check now. Note that from (4.4) the vector of the parameters is $\theta=(A'_{11}, A'_{22}, \eta_{11}, \eta_{22})^T$ and $\theta^0=(A'^0_{11}, A'^0_{22}, \eta^0_{11}, \eta^0_{22})^T$ is its true value. Note that we assume again that Φ is of the form (B.9). Finally, we write s_{11}, s_{12} and s_{22} the entries of the matrix Σ (without loss of generality, we could as well take Σ the identity matrix, as was done for the proof of the previous example).

B.2.1 Assumption $\mathbf{H}_{3.6}$

We start by making an important observation, namely that the matrix $V=(V_{i,j})_{1\leq i,j\leq 4}$ is block-diagonal, with blocks for $(i,j)\in\{1,2\}$ and $(i,j)\in\{3,4\}$, respectively.

Regarding the second block, tedious calculations (therefore carried out with Mathematica) for $\theta_3=\eta_{11}$ yield

$$\begin{aligned}&\operatorname{tr}\left[\left\{\Sigma_t^{-1}(\theta)\frac{\partial\Sigma_t(\theta)}{\partial\theta_3}\Sigma_t^{-1}(\theta)\frac{\partial\Sigma_t(\theta)}{\partial\theta_3}\right\}_{\theta=\theta^0}\right]\\ &\qquad = 2\sin^2(ct)\frac{\left(\left(e^{\eta_{22}\sin(ct)}s_{11}-s_{12}\right)^2+2(s_{11}s_{22}-s_{12}^2)\right)}{\left(1+e^{(\eta_{11}+\eta_{22})\sin(ct)}\right)^2(s_{11}s_{22}-s_{12}^2)}=V_{33}(t),\end{aligned}$$

where $V_{ij}(t)$ is the term t in the sum defining V_{ij}, $i,j=1,2,3,4$. Since Σ is an invertible matrix, $s_{11}s_{22}-s_{12}^2=\det(\Sigma)>0$, hence this expression is clearly positive. Bounding the numerator is straightforward (since $|\sin(ct)|<1$), and $1/\left(1+e^{(\eta_{11}+\eta_{22})\sin(ct)}\right)^2$ can simply be bounded by 1. Hence, there exist a constant α not depending on t such

that $V_{33}(t) \leq \alpha$, consequently the entry V_{33} is positive and finite. The same conclusion obviously holds for V_{44}. Turning our attention towards V_{34}, we obtain

$$\begin{aligned}
&\operatorname{tr}\left[\left\{\Sigma_t^{-1}(\theta)\frac{\partial \Sigma_t(\theta)}{\partial \theta_3}\Sigma_t^{-1}(\theta)\frac{\partial \Sigma_t(\theta)}{\partial \theta_4}\right\}_{\theta=\theta^0}\right] \\
= 2\sin^2(ct)&\frac{\left(s_{12}\left(e^{\eta_{22}\sin(ct)}s_{11} - s_{12} - e^{\eta_{11}\sin(ct)}s_{22}\right) - e^{(\eta_{11}+\eta_{22})\sin(ct)}(s_{11}s_{22} - 2s_{12}^2)\right)}{\left(1 + e^{(\eta_{11}+\eta_{22})\sin(ct)}\right)^2\left(s_{11}s_{22} - s_{12}^2\right)} = V_{34}(t).
\end{aligned}$$

Again, it is a simple exercise to show that this term is bounded independently of t.

Let us turn our attention towards the block $(1,2)$ now. From (3.7) we know that

$$\frac{\partial e_t^T(\theta)}{\partial \theta_i} = \sum_{k=1}^{t-1}\psi_{tik}(\theta,\theta^0)g_{t-k}\epsilon_{t-k}.$$

Following Section B.1.6 for $i,j=1$ with $\theta_1 = A'_{11}$, we obtain a generalized version of (B.15) as

$$\begin{aligned}
E_{\theta^0}\left(\frac{\partial e_t^T(\theta)}{\partial \theta_1}\Sigma_t^{-1}\frac{\partial e_t(\theta)}{\partial \theta_1}\right) &= E_{\theta^0}\left[\left(\sum_{k=1}^{t-1}\psi_{t1k}(\theta,\theta^0)g_{t-k}\epsilon_{t-k}\right)^T\Sigma_t^{-1}\sum_{k=1}^{t-1}\psi_{t1k}(\theta,\theta^0)g_{t-k}\epsilon_{t-k}\right] \\
&= \sin^2(at)\sum_{k=1}^{t-1}\operatorname{tr}\left[\begin{pmatrix} A_{t,1,1}^{(k-1)} & A_{t,1,2}^{(k-1)} \\ 0 & 0 \end{pmatrix}\Sigma_{t-k}\begin{pmatrix} A_{t,1,1}^{(k-1)} & A_{t,1,2}^{(k-1)} \\ 0 & 0 \end{pmatrix}^T\Sigma_t^{-1}\right],
\end{aligned}$$

an expression clearly bounded (under the same conditions as those in B.1). Its positiveness follows from the fact that we consider a Mahalanobis distance in the metric Σ_t^{-1}.

Finally, checking that the blocks $(1,2)$ and $(3,4)$ are positive definite has been done numerically in the numerical example of Section 5.2.

B.2.2 Assumption $\mathbf{H}_{3.7}$

In the first part of this assumption we have to show that

$$\frac{1}{n^2}\sum_{d=1}^{n-1}\sum_{t=1}^{n-d}\sum_{k=1}^{t-1}\|g_{t-k}\|_F^2\left\|\psi_{tik}^{(n)}\right\|_F\left\|\psi_{t+d,i,k+d}^{(n)}\right\|_F = O\left(\frac{1}{n}\right).$$

From (4.6) we have

$$\|g_{t-k}\|_F^2 = 2 + e^{-2\eta_{11}\sin(ct)} + e^{-2\eta_{22}\sin(ct)}$$

which is an easily bounded function. Consequently, we are left with exactly the same expression as for Example 4.1, which solves the question.